A little knowledge

The Japanese authorities have done well in releasing copious amounts of crude data on the nuclear crisis. But it is imperative for the data to be provided in more meaningful and user–friendly ways.

One month after the earthquake and tsunami hit Japan, there is still no clear picture of the further hazard posed by the wrecked nuclear reactors and spent fuel ponds at the Fukushima nuclear power plant (see page 146), and monitoring of fallout remains patchy. To improve the situation, better data, in more user-friendly forms, and more sophisticated analyses are essential. Compared with the 1979 Three Mile Island accident or the 1986 Chernobyl disaster, there is certainly much more information available about this latest nuclear accident — largely thanks to the Internet and online media. Japan's science ministry, and other bodies, have issued reams of data, including daily environmental radiation measurements — an admirable feat, given that the Japanese authorities are also having to deal with the huge aftermath of the quake and tsunami.

But as Peter Sandman, a risk consultant based in Princeton, New Jersey (www.psandman.com/whatsnew.htm), points out, the authorities have failed badly to forewarn the public of a series of events that they must have known were likely to happen. This has resulted in nasty surprises such as radioactive pollution of the sea (see page 145), foods and tap water — as well as this week's upgrading of the accident to level 7, the highest on the International Nuclear and Radiological Event Scale and matched only by Chernobyl. As a result, many people now do not trust the authorities to tell them if the situation is likely to worsen, sapping public confidence.

The Tokyo Electric Power Company (TEPCO), which runs the Fukushima plant, has also on at least four occasions had to retract as incorrect its findings on the amount and composition of radionuclides in areas in and around the plant, or on reactor parameters. This has created uncertainty and public mistrust in the company's monitoring abilities. In its defence, damaged plant instrumentation means that key data on events inside the reactors are sometimes missing or unreliable. Even so, the most complete and credible publicly available analyses of possible reactor-event scenarios have come not from Japan, but from outside scientists, nuclear-reactor makers and regulatory authorities.

Similarly, it is pertinent to ask why, so far, the only detailed publicly available forecasts of the direction and concentration of atmospheric radionuclide plumes have come from overseas agencies. The Japanese almost certainly have data that would allow much higher-resolution forecast maps of Fukushima and the surrounding areas. Although the Japanese authorities are releasing data daily on radiation levels in the air, soil and water, these are scattered across multiple, individual web pages. This uncoordinated approach was excusable in the early days, but data collection and presentation urgently need to improve. The authorities have also failed to provide vital context on how these exposure rates translate (or not) into what matters to people, such as health effects, and where they make farming impossible. The recurring narrative that this or that radiation dose is as much as would be given by an X-ray or a CT scan doesn't cut it, as health effects, for example, depend most on accumulated doses over long times.

Information on fallout distribution made public by the government and TEPCO also lacks basic metadata, such as the latitude and longitude of sampling points or the sampling protocols used, and results are presented as static PDFs from which researchers cannot easily extract the data. As a result, it is next to impossible for academic researchers and others to compile and map the daily reports and gain a better picture of the situation and of changes over time and space. The Japanese authorities, the International Atomic Energy Agency and other bodies with relevant information must present it as dynamic data and high-resolution maps that also show day-to-day variation, total net cumulative soil deposition and where hotspots are, and as models of what's happening overall rather than just spot counts.

> *"The authorities have failed badly to forewarn the public of events they must have known were likely to happen."*

Data analysis should also not be left to governments alone. Researchers are rightly calling for an independent group to process the data and publish evidence-based risk assessments. They also want data in machine-readable formats, such as spreadsheets, databases and spatial data formats. This would unleash the diverse creativity of academic researchers, journalists, software geeks and mappers, who are often better equipped, and more agile than governments and international agencies, to present data online in timely, informative and compelling ways. To convert raw data into high-quality, user-friendly forms is not a luxury, but essential for helping to build public trust. ∎

This is not a drill

Worst–case scenarios happen. We must prepare for them and be able to respond.

Natural disasters and man-made tragedies have always been a feature of life on Earth, but the world does seem to have become particularly hazardous of late.

A year ago, an explosion on a BP drilling rig in the Gulf of Mexico killed 11 workers and triggered a leak that spewed more than 4 million barrels of oil into the ocean. As we report on page 152, scientists are still struggling to track the possible long-term effects on the region. Meanwhile, halfway across the world, Japan is reeling from the triple crisis that killed tens of thousands, caused at least US$200 billion in direct damages and continues to spread radiation into the skies and sea.

The two disasters are of different magnitude and have very different causes. But they do have one thing in common: overconfidence in the strength of human systems and decisions.

In the case of the petroleum industry, companies pushed into deep-water locations without doing sufficient research or making the investments needed to prevent — and respond to — problems that might arise from drilling in such an extreme environment. And the US government did not have sufficient oversight of the industry. Both the government and the petroleum companies seemed to think that there was little risk of the type of blowout that caused the accident. "The Deepwater Horizon disaster exhibits the costs of a culture of complacency," concluded the presidential commission charged with investigating the spill.

Similar overconfidence ruled parts of the seismological community in Japan. Official maps of the seismic hazard in the country draw heavily on records of past earthquakes. But because no great earthquake had struck off the coast of Sendai in recent centuries, the hazard assessments did not take such a large event into account (see *Nature* doi:10.1038/nature10105; 2011, and *Nature* **471**, 556–557; 2011). Some researchers were more cautious. They pointed to geodetic data showing that strain was increasing in the region and to geological signs of a tsunami in 869 that was much larger than anything more recent. But such evidence did not undermine seductive faith in the official quake risk-assessment method, which turned out to be fundamentally flawed. In addition, the designers of the Fukushima nuclear power plant failed to adequately prepare for the possibility that their back-up generators would fail, and misjudged how hard it would be to re-establish electrical power after a tsunami.

One lesson that must be taken from these tragic events is that many of our critical systems are simply unable to withstand situations that are entirely possible. Think the oil and gas industry has learned the lessons of Deepwater Horizon? Don't hold your breath. The rush to exploit the resources exposed by dwindling Arctic ice cover seems riddled with the same dismissal of legitimate risks of rare but plausible events with terrible consequences (see page 162).

Such critical systems — and society at large — must be made more resilient: the core elements of society need to function even when disasters strike. And many fields, ranging from seismic engineering to urban planning for the impacts of climate change, are working to build up this resilience.

"Critical systems must be made more resilient."

In practice, this means strengthening crucial buildings and other infrastructure, developing back-up power systems and planning for multiple tiers of failures. It also requires better training for local communities and government officials in how to respond. Earthquake and tsunami drills are common in places such as California and Japan, but other regions, too, need to engage in these and other types of exercises that simulate massive oil spills, nuclear crises, terrorist attacks and hurricanes, to name a few.

On 28 April, almost two million people in the United States will take part in the Great Central US Shakeout by responding as if a large earthquake had taken place in the New Madrid Seismic Zone, which 200 years ago produced the strongest historical quakes in the conterminous United States. This relatively simple public exercise will be followed up a few weeks later by a more detailed drill involving federal, state and local managers.

Such simulations are essential to probe emergency plans for weaknesses before the real hazards come along, which they will. ∎

Universal truths

Rejection of broad commonality in structure of languages has implications for all sciences.

Since at least the days of Aristotle, a search for universal principles has characterized the scientific enterprise. In some ways, this quest for commonalities defines science: without it, there is no underlying order and pattern, merely as many explanations as there are things in the world. Newton's laws of motion, the oxygen theory of combustion and Darwinian evolution each bind a host of different phenomena into a single explicatory framework.

In physics, one approach takes this impulse for unification to its extreme, and seeks a theory of everything — a single generative equation for all we see. It is becoming less clear, however, that such a theory would be a simplification, given the proliferation of dimensions and universes that it might entail. Nonetheless, unification of sorts remains a major goal.

This tendency in the natural sciences has long been evident in the social sciences too. Here, Darwinism seems to offer justification, for if all humans share common origins, it seems reasonable to suppose that cultural diversity could also be traced to more constrained beginnings. Just as the bewildering variety of human courtship rituals might all be considered to be forms of sexual selection, perhaps the world's languages, music, social and religious customs and even history are governed by universal features. To filter out what is contingent and unique from what is shared might enable us to understand how complex cultural behaviour arose and what guides it in evolutionary or cognitive terms.

That, at least, is the hope. But a comparative study of linguistic traits published online today (M. Dunn *et al. Nature* doi:10.1038/nature09923; 2011) supplies a reality check. Russell Gray at the University of Auckland, New Zealand, and his colleagues consider the evolution of grammars in the light of two previous attempts to find universality in language.

The most famous of these efforts was initiated by Noam Chomsky, who postulated that humans are born with an innate language-acquisition capacity — a brain module or modules specialized for language — that dictates a universal grammar. A few generative rules are then sufficient to unfold the entire fundamental structure of a language, which is why children can learn it so quickly. Languages would diversify through changes to the 'parameter settings' of the generative rules.

The second, by Joshua Greenberg, takes a more empirical approach to universality, identifying traits (particularly in word order) shared by many languages, which are considered to represent biases that result from cognitive constraints. Chomsky's and Greenberg's are not the only theories on the table for how languages evolve, but they make the strongest predictions about universals.

Gray and his colleagues have put them to the test using phylogenetic methods to examine four family trees that between them represent more than 2,000 languages. A generative grammar should show patterns of language change that are independent of the family tree or the pathway tracked through it, whereas Greenbergian universality predicts strong co-dependencies between particular types of word-order relations (and not others). Neither of these patterns is borne out by the analysis, suggesting that the structures of the languages are lineage-specific and not governed by universals.

This does not mean that cognitive constraints are irrelevant, or that there are no other universals dictated by communication efficiency. It is surely inevitable that cognition sets limits on, say, word length. But such 'universals' seem likely to be relatively trivial features of languages, just as may be the case for putative universals in music and other aspects of culture.

The conclusion? We should perhaps learn the lesson of Darwinism: a 'universal' mechanism of adaptation says little in itself about how a particular feature got to be the way it is, or about how it works. This truth has dawned on physicists too: universal equations are all very well, but the world actually consists of particular solutions, and these are generally the result of contingent history. One size does not always fit all. ∎

↪ **NATURE.COM**
To comment online, click on Editorials at:
go.nature.com/xhunqv

M. HOFFERT

Governments must pay for clean-energy innovation

*The current obsession with nuclear power is a red herring, says **Marty Hoffert**. The United States and others should instead invest in a clean-energy revolution.*

There has been much debate about the future of nuclear power following the crisis in Japan. Yet present-day fission reactors are, at best, transition technology. Instead, we should use the current high profile of energy to refocus the debate on what the long-term solution to our problems must be: innovation.

As a scientist working on energy and climate change, I was inspired by President Barack Obama's call for US researchers to put a million electric cars on the road by 2015, and to generate 80% of our electricity from carbon-neutral sources by 2035. In this journal in 1998, colleagues and I concluded that engineering projects on the scale of the Apollo moonshots might be needed to transform the world's energy system (M. I. Hoffert *et al. Nature* **395,** 881–884; 1998). We should now be debating not whether but how to do this.

Some assert that government investment in transformative energy technology is code for tax and spend, and that suitable technologies already exist, or will be delivered by market forces. Others argue that government incentives such as feed-in tariffs for solar and wind energy are unnecessary, and that clean energy should compete in the market from the beginning. But the idea that private-sector entrepreneurship can do the job alone is based on a myth. It took 30 years of government funding of the Internet by the military research agency DARPA and the National Science Foundation before Wall Street discovered that there was money to be made out of it.

The private-sector-alone approach is a prescription for disaster, and displays abysmal ignorance of how the United States ended up with its current energy system. The US government made crucial investments in energy technology in the post-war years. Consider Hyman Rickover's light-water reactor, developed for *Nautilus,* the world's first nuclear submarine, which became the prototype for 85% of the world's nuclear power plants. We still use the term 'reactor fleet'.

Today, a poster child for carbon capture and storage (CCS) schemes is carbon dioxide collection at the Dakota Gasification Company's plant in Beulah, North Dakota. This facility was paid for initially by the US Department of Energy as a synfuel plant under President Jimmy Carter's Energy Independence Program. It would not have happened without massive government underwriting of the risk. CCS in the North Sea by the Norwegian Statoil Company is likewise heavily state subsidized.

No money? US debt now is a comparable per cent of gross domestic product (GDP) to what it was in the Great Depression before the Second World War. By massively borrowing from ourselves to finance President Franklin D.

Roosevelt's miracles of war production and technology development, we saw aircraft morph from biplanes to jets, and nuclear power become a reality, even as the US debt-to-GDP ratio increased to more than 100% by the end of the war. We bet the farm on a stimulus package on steroids — and we won. The United States emerged as the strongest economy on the planet.

One can only hope that we're not so distracted by ideological battles about government versus private-sector funding that the real energy and global-change problems defeat us because of a failure of imagination — particularly a failure to fund research, development and demonstration in sustainable energy, at least with the initial US$15 billion a year recommended by industry leaders and academic researchers. If the president asked Congress for this sum, which could even be paid for by eliminating perverse subsidies for fossil fuels, it would still be only one-tenth of 1% of the present US GDP of $15 trillion. We can afford the investment needed to induce a revolutionary transformation of the world energy system away from fossil fuels. China is planning to invest $75 billion a year to do just that.

> IN THE LATTER PART OF THE 'AMERICAN CENTURY', THE **UNITED STATES** SOMEHOW **LOST ITS WAY.**

In the latter part of the 'American century', the United States somehow lost its way. No longer 'makers', we became a nation of rustbelts, Ponzi schemes and subprime mortgage risk, myopically focused on quarterly earnings and consumerism. What a tragedy it would be to lose America's talent for innovation after 200 years.

Mr President and Congress: open your minds to a civilization powered by wind turbines in harmony with our landscape and continental shelves; solar electricity from deserts and Earth orbit powering our cites; safe, proliferation-resistant nuclear reactors; coal gasifiers driving efficient electric power plants with CO_2 stored underground; along with energy-efficient homes and public buildings, smart power grids, high-speed rail, electric and biofuelled cars, even carbon-neutral fuels made from sunlight, water and CO_2 in the atmosphere more efficiently than nature does by photosynthesis. These are no longer impossible dreams, but realities of new US industries revitalized by American entrepreneurs and a high-tech workforce, much like the one Roosevelt created to fight the Second World War.

This is a dream worth rededicating the American experiment to: visionary, and yet science-based, that goal will lift the spirit of our children and grandchildren with passion and the tenacity to make it so. Say it, Barack, shout it from the rooftops, dedicate your presidency to it, and you will stand immortal in the pantheon of American leaders who changed everything. ∎

Marty Hoffert *is professor emeritus of physics at New York University.*
e-mail: marty.hoffert@nyu.edu

Map of missing and extra DNA

A study has unearthed nearly 2 million instances of additional and missing stretches of DNA from the genomes of 79 people. The analysis of archived genome sequences revealed insertions and deletions (indels) ranging from 1 to 10,000 base pairs in length. Almost two-thirds have not been reported before.
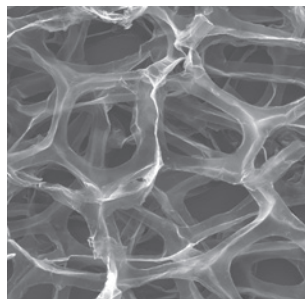
Researchers have struggled to identify such small, but important, variations in the human genome. More than 2,000 of the indels found by Scott Devine at the University of Maryland in Baltimore and his team were in the coding regions of known genes. The authors developed a chip-based assay for quick identification of about 10,000 of the smaller indels. Such chips could aid in interpreting the results of other genomic studies.

Genome Res. doi:10.1101/gr.115907.110 (2011)

Fabricating a graphene foam

Graphene has remarkable electronic properties, but using these atom-thin sheets of carbon in devices remains challenging, and attempts to build graphene macrostructures have

A. DARRINGTON/ALAMY

Hidden in plain sight

Some species elude predators by masquerading as common, inedible objects. But far from being a passive disguise, such camouflage also involves specific changes in prey and predator behaviour, say John Skelhorn at the University of Exeter, UK, and his colleagues.

In lab experiments, the researchers exposed twig-like *Selenia dentaria* caterpillars (**pictured**) sitting among real twigs to domestic chicks trained to attack them. Not only did the chicks take longer to spot their prey when twig density was higher, but caterpillars also preferred branches with more twigs. During the day, caterpillars opted for twig density over leaf availability; however, at night, when predators are no longer on the prowl, the insects were found on branches with abundant leaves to eat.

High twig density further protects caterpillars by dashing predators' hopes. Repeated exposure to branches with many twigs and no caterpillars made chicks less motivated to hunt for prey than birds less frequently exposed to this situation.

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.1014629108 (2011)

yielded materials with low conductivity. Hui-Ming Cheng and his co-workers at the Chinese Academy of Sciences in Shenyang now report a three-dimensional, conductive all-graphene macrostructure they call a graphene foam.

The researchers deposited graphene on a template made from a porous nickel foam. After removing the nickel, they were left with a structure made of a three-dimensional network of interconnected graphene channels (micrograph **pictured**). Combined with a silicon-based polymeric matrix, the foam forms a composite that retains the high conductivity and flexibility of the two components, making this a promising material for flexible and large-scale electronic applications.

Nature Mater. doi:10.1038/nmat3001 (2011)

Animals go 'blue' with temperature

An analysis of more than three decades' worth of field data reveals that the rate at which environmental temperatures in a given area oscillate is closely matched by rates of oscillations in the size of resident animal populations.

Population biologists use colours to describe these changes. Species whose numbers oscillate mainly over the long term are called red, whereas those shifting over shorter timescales are termed blue. Bernardo García-Carreras and Daniel Reuman at Imperial College London analysed data on 147 species and records from weather stations around the world. They show that environments seem to tinge local populations the same colour. Furthermore,

environments on most continents have grown bluer with climate change, so animal populations may also become bluer.

Previous models have suggested that those red populations that shift towards blue are less prone to extinction, although other factors, such as habitat destruction, could counteract any benefit.
J. Anim. Ecol. doi:10.1111/j.1365-2656.2011.01833.x (2011)

Birds with beauty — and brains

Some birds have it all. Male siskins with brighter, more attractive plumage are better at finding food than their less eye-catching counterparts.

Fernando Mateos-Gonzalez and his team at the Natural History Museum of Barcelona, Spain, measured the length of a yellow stripe on the wings of male siskins (*Carduelis spinus*; **pictured**). Females are known to pick mates on the basis of the length of this stripe. The researchers challenged hungry males to unlock a cache of pine seeds blocked by toothpicks.

Birds with longer yellow stripes solved the problem more quickly than those with shorter stripes. The authors ruled out age, size and social dominance as possible contributory factors, and conclude that fancier feathers might help discerning females to select mates that are not only sexier, but also smarter.
Biol. Lett. doi:10.1098/rsbl.2011.0163 (2011)

Saturated fats up inflammation

A diet high in saturated fats raises the risk of type 2 diabetes — perhaps owing to the activity of an inflammatory protein complex called the inflammasome.

Jenny Ting and her team at the University of North Carolina at Chapel Hill found that a saturated fatty acid called palmitate stimulates the inflammasome — which activates the inflammatory response — in cultured mouse macrophages, a type of immune cell. An unsaturated fat did not do this. The macrophages, in turn, produced a potent inflammatory molecule called IL-1β that blocked insulin signalling in cultured liver cells. Macrophages lacking key components of the inflammasome did not interfere with insulin signalling.

Mice fed a high-fat diet for 12 weeks developed insulin resistance in liver, fat and muscle tissue, along with other signs of type 2 diabetes, but those that lacked inflammasome genes did not.
Nature Immunol. doi:10.1038/ni.2022 (2011)

Stem-cell genes linked to relapse

The recurrence of colon cancer in patients after treatment is linked to a pattern of genes expressed in the tumour that is specific to intestinal stem cells.

Eduard Batlle at the Institute for Research in Biomedicine in Barcelona, Spain, and his colleagues isolated cells from pockets in the mouse intestinal wall known to house stem cells. They found that intestinal stem cells had the highest expression of a gene called *Ephb2*, and that about 50 other genes were also upregulated. Analysis of samples from 340 patients with colon cancer revealed an association between high expression levels of these genes and increased risk of relapse.

When human tumour cells were implanted in mice, cells expressing the highest levels of *Ephb2* showed the greatest capacity for initiating tumours at low cell dosages.
Cell Stem Cell doi:10.1016/j.stem.2011.02.020 (2011)

Cold spells in a warm world

Severe cold streaks could persist well into the future despite global warming, a climate modelling study finds.

Auroop Ganguly and his colleagues at Oak Ridge National Laboratory in Tennessee compared data from extreme cold events between 1991 and 2000 with projections made by nine global climate models for 2091–2100. Their results show that by the end of this century, extreme cold snaps in many parts of the world could be more intense and more prolonged — although less frequent — than the twentieth-century average.

The team suggests that governments will still need to plan for cold spells even as the world warms.
Geophys. Res. Lett. doi:10.1029/2011GL047103 (2011)

Laser from a tiny wire

⭐ **HIGHLY READ** on pubs.acs.org in March

Lasers made from nanometre-scale wires could be valuable for small, fast, optical-communication and signal-processing devices, but researchers have been unable to tune the wavelength of the light emitted.

Limin Tong at Zhejiang University in Hangzhou, China, and his colleagues have made a laser from a cadmium selenide wire 200 nanometres in diameter and bent into two loops — resembling an elongated figure of eight. The loops act as mirrors to reflect the light back and forth, boosting the light's intensity. When the researchers decreased the size of one of the loops, the peak wavelength of the light shifted by 7 nanometres, suggesting that lasers created in this way could be made fully tunable.
Nano Lett. 11, 1122–1126 (2011)

No DNA needed, RNA goes solo

Researchers have engineered an RNA enzyme to synthesize another active RNA enzyme from an RNA template. This models one theory for how life originated on Earth: with RNA molecules that both encoded genetic information and catalysed reactions to express that information.

Philipp Holliger and his co-workers at the Medical Research Council Laboratory of Molecular Biology in Cambridge, UK, started out with an RNA enzyme, or ribozyme, called R18, which is known to synthesize short stretches of RNA. They created a library of modified versions of R18 and used it to screen for ribozymes with improved enzymatic activity. They ended up with a new ribozyme, tC19Z, capable of generating longer pieces of RNA from a wider range of templates than R18. The enzyme could also synthesize a different ribozyme.
Science 332, 209–212 (2011)

↻ **NATURE.COM**
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

EVENTS

Japan disaster

Japan's nuclear regulator now rates the accident at the Fukushima power plant at 'level 7' — the highest possible on a safety scale set by the International Atomic Energy Agency, and only previously used to describe the Chernobyl disaster in 1986. The Japanese Nuclear and Industrial Safety Agency upped its assessment by two levels on 12 April, after re-evaluating the amount of radiation released, which it said was one-tenth of that from Chernobyl. Japan's government is also expanding the evacuation zone around the plant to some regions beyond the current 20-kilometre radius. Meanwhile, the country continues to be rocked by strong aftershocks. Besides a constant stream of magnitude 4 and 5 earthquakes, a 7.1-magnitude quake struck on 7 April; and one of magnitude 6.6 hit on 11 April — a month after March's 9.0-magnitude quake. For more on Fukushima, see pages 143–146 and www.nature.com/japanquake.

Eleven thousand metres under the sea

British billionaire entrepreneur Richard Branson last week launched plans to send a manned submarine on five dives to the deepest parts of the oceans. The Virgin Oceanic mission is set to start in late 2011 at the 11-kilometre-deep Mariana Trench in the Pacific Ocean. It would be the first manned exploration of the area since the *Trieste* submersible visited the Challenger Deep area of the trench in 1960. A series of remotely controlled probes with research equipment will also be sent into the depths. See go.nature.com/rgy9za for an interview with scientists on the team.

POLICY

US budget cuts

Scientists at US federal agencies avoided the prospect of a government shutdown when congressional leaders agreed — barely an hour before the deadline on 8 April — on an outline budget to fund the government until the end of the current fiscal year on 30 September. That deal slashes around US$38.5 billion off 2010 spending levels, but as *Nature* went to press details of how the cuts would affect researchers had not been negotiated. Despite earlier Republican calls for 5–18% cuts to the National Science Foundation and the Department of Energy's Office of Science, a White House statement said that those agencies would still see strong investments.

Genetic screening

Genetic tests for a range of inheritable conditions should be widely available to potential parents, according to a UK government advisory group, the Human Genetics Commission. Screening is already available for people with a high chance of passing on some genetic disorders, such as cystic fibrosis or Tay-Sachs disease. But the commission's report, released on 6 April, found "no specific ethical, legal or social principles" that would make wider genetic screening before conception unacceptable. The country's National Screening Committee is now considering the report.

Resistant bacteria

India's government has reacted strongly against the discovery of antibiotic-resistant bacteria in drinking water and sewage from New Delhi, calling it "unsupported by any clinical or epidemiological evidence". Researchers reported identifying bacteria carrying a gene (bla_{NDM-1}) that confers resistance to a wide range of antibiotics. Although NDM-1-positive strains have been seen in patients in many countries, the study was the first to find them in environmental samples unconnected with hospital settings (T. R. Walsh *et al.* *Lancet Infect. Dis.* doi:10.1016/S1473-3099(11)70059-7; 2011). Its publication coincided with World Health Day (7 April), on which the World Health Organization called for more efforts to prevent the spread of antibiotic resistance. See go.nature.com/2qmhec for more.

Ethical biofuels

Certification standards are needed to ensure that biofuels are produced ethically and sustainably, says a report released on 13 April by the Nuffield Council on Bioethics, an independent charity based in London that studies ethical issues in research. Referring to European policies, the council says that targets encouraging

rapid uptake of biofuels neglect wider ethical concerns, such as deforestation and the displacement of indigenous people. The report suggests that biofuels development should avoid human-rights violations, adhere to fair trade principles, reduce greenhouse-gas emissions, and take into account the needs of both local and international communities.

Australian protests

On 12 April, thousands of researchers rallied in Melbourne, Sydney, Canberra and Adelaide to protest against rumoured cuts in government funding for medical research. The National Health and Medical Research Council may lose Aus$400 million (US$419 million) over the next three or four years, if leaked figures are accurate. That amounts to a cut of 14–19% from the council's Aus$715 million 2010–11 budget. A fifth rally was planned for Perth on 14 April; the budget is due on 10 May. See go.nature.com/8lz2if for more.

EPA skirmish

The US Environmental Protection Agency (EPA) has survived political attempts to block its regulation of greenhouse-gas emissions. On 6 April, the US Senate rejected an amendment to a business bill that would have eliminated the EPA's authority to act on greenhouse gases — as well as three minor amendments that would each have restricted the agency. This neutralized the passing of similar legislation on 7 April, by the Republican-controlled House of Representatives. But the struggle is far from over: having failed to curtail the EPA's authority, Republicans are now looking to cut the agency's funding as part of broader budget wrangles.

Suicides at KAIST

The suicides of four students since the start of the year at the prestigious Korea Advanced Institute of Science and Technology (KAIST) in Daejeon are putting pressure on its president, Nam Pyo Suh. An assistant director at the US National Science Foundation in 1984–88 and head of mechanical engineering at the Massachusetts Institute of Technology from 1991 to 2001, Suh came to the state-financed KAIST in South Korea in 2006 to improve competition. But after the most recent suicide, on 7 April, Korean media reported faculty and students criticizing his educational reforms. Suh has reportedly agreed to scrap a policy that links student fees to academic performance.
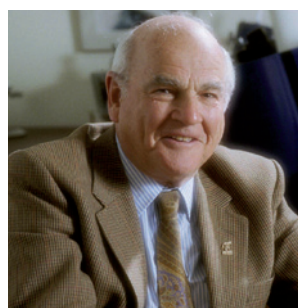
Hep B expert dies

Baruch Blumberg, who shared the 1976 Nobel Prize in Physiology or Medicine for his research on infectious disease, died on 5 April, aged 85. Blumberg (**pictured**) identified the hepatitis B virus, and showed how it causes liver disease. He spent much of his career in Philadelphia, at the Fox Chase Cancer Center, and later served as the first director of NASA's Astrobiology Institute in Moffett Field, California.

Nitrogen warning

Overuse of nitrogen fertilizers costs the European Union between €70 billion (US$100 billion) and €320 billion per year, according to a landmark analysis of nitrogen flows across Europe, released on 11 April. Some 200 researchers tallied up the costs and benefits of producing compounds that contain nitrogen, such as ammonia and nitrogen oxides. They concluded that, financially, the detrimental effects on human health and biodiversity far outweighed the benefits for crop production. A worldwide assessment and agreements to reduce fertilizer production are now needed, they say. See page 159 for more.

18–19 APRIL
Some details about the future of NASA — including updates on missions with the European Space Agency — might be revealed at a public meeting of its science advisory council's planetary science subcommittee in Washington DC.
go.nature.com/oxfmvk

18–19 APRIL
The Royal Society in London holds a meeting on the best ways to model ecological systems to predict their futures.
go.nature.com/yjyz5u

Biotech pain

All five biopharmaceutical firms to enter public markets this year in the United States have had to drastically cut the price at which they offer shares to attract investors — a sign of weak demand in the sector. On 5 April, stem-cell therapy company Aldagen, based in Durham, North Carolina, said it was withdrawing its plans for an initial public offering (IPO), citing market conditions. A day earlier, Tranzyme, also based in Durham, began trading after cutting prices three-fold to $4 a share. IPOs in other fields have not had to cut share pricings.

↪ **NATURE.COM**
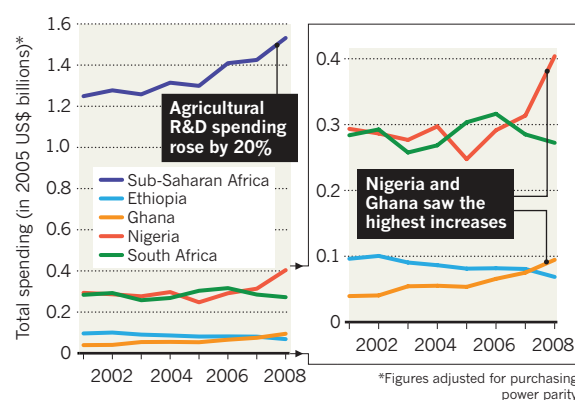For daily news updates see:
www.nature.com/news

TREND WATCH

"Progress for some; challenges for many," finds a report analysing spending in agricultural research and development (R&D) in sub-Saharan Africa. After a decade of stagnation in the 1990s, spending rose by more than 20% from 2001 to 2008. But most of the growth occurred in a handful of countries. In many others, particularly in francophone West Africa, spending declined. The report urges researchers not to rely on temporary loans or unstable inflows of donor funding. See go.nature.com/dxypxv for more.

AFRICA'S AGRICULTURAL RESEARCH
Countries have benefited unevenly from sub-Saharan Africa's rise in spending on agricultural research (2001–08).

Agricultural R&D spending rose by 20%

Sub-Saharan Africa
Ethiopia
Ghana
Nigeria
South Africa

Nigeria and Ghana saw the highest increases

Total spending (in 2005 US$ billions)*

*Figures adjusted for purchasing power parity

NEWS IN FOCUS

With nuclear power stations out of action, can Japan find a way to reduce carbon emissions and still avoid the blackouts?

ENERGY

Japan faces power struggle

Climate goals in doubt as fossil fuels come to the fore in fight to meet electricity demand.

BY JEFF TOLLEFSON

The disaster at Japan's Fukushima nuclear power plant is putting the nation's ambitious plans to reduce carbon emissions under serious pressure. With few natural resources, Japan has long been forced to rely on imports of oil, coal and natural gas to provide about 80% of the energy needed to sustain its economy. Nuclear energy was supposed to help Japan limit these energy imports and cut its carbon emissions at the same time — but that was before Fukushima.

The government has committed to reduce carbon emissions by 25% from 1990 levels by 2020, and by roughly 80% from 1990 levels by 2050. The short-term target is particularly ambitious, because Japan's streamlined economy is already among the most energy-efficient in the world. To achieve its goals, Japan had planned to build nine nuclear power plants over the coming decade and a total of 14 plants by 2030, all while squeezing as much power as possible out of the current 54 plants. By 2030, nuclear power was to have provided half of Japan's electricity, doubling its current contribution (see 'Nuclear power in Japan').

A big question mark now hangs over those plans. Nearly a month after the magnitude-9.0 earthquake and the tsunami that followed, workers are still struggling to control the reactors at Fukushima. Four

![FUKUSHIMA CRISIS]

➲ **WWW.NATURE. COM/JAPANQUAKE**

of the six reactors there have been destroyed; the other two will probably never run again; and plans for two new reactors have been abandoned (see page 146). Another eight reactors in the northwest of the country were automatically shut down after the 11 March earthquake. The Tokyo Electric Power Company (TEPCO), Japan's largest private utility, which operates Fukushima Daiichi and three other nuclear plants with multiple reactors, is struggling to meet electricity demand in Tokyo. Japan's vulnerability was highlighted again on 7 April, when a magnitude-7.1 earthquake temporarily shut down several conventional power stations and left millions without electricity.

No one knows when the situation will stabilize, nor what Japan's nuclear industry will look like when it does. Although several power ▶
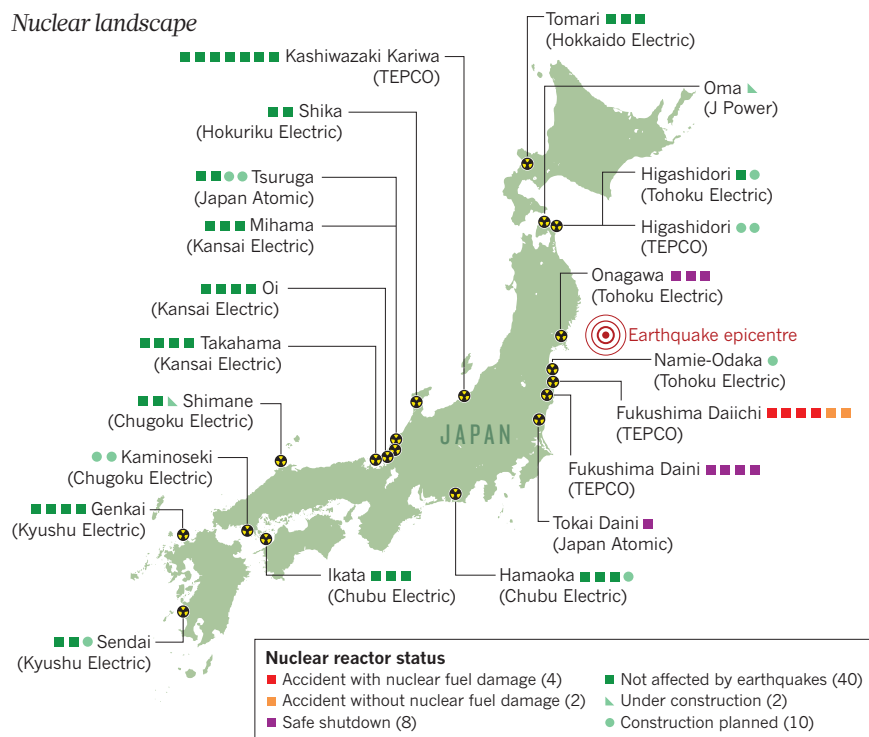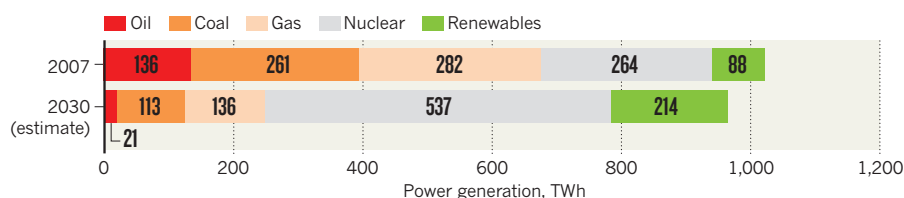
SOURCE: METI/JAIF

NUCLEAR POWER IN JAPAN

Fifty-four nuclear reactors supply about a quarter of Japan's electrical energy needs (**1**). Before the Fukushima disaster struck, new reactors and energy conservation were set to double that proportion by 2030 (**2**).

1 Nuclear landscape

Tomari (Hokkaido Electric)

Oma (J Power)

Kashiwazaki Kariwa (TEPCO)

Shika (Hokuriku Electric)

Higashidori (Tohoku Electric)

Higashidori (TEPCO)

Tsuruga (Japan Atomic)

Mihama (Kansai Electric)

Onagawa (Tohoku Electric)

Oi (Kansai Electric)

Earthquake epicentre

Takahama (Kansai Electric)

Namie-Odaka (Tohoku Electric)

Shimane (Chugoku Electric)

Fukushima Daiichi (TEPCO)

Kaminoseki (Chugoku Electric)

Fukushima Daini (TEPCO)

Genkai (Kyushu Electric)

Tokai Daini (Japan Atomic)

JAPAN

Ikata (Chubu Electric)

Hamaoka (Chubu Electric)

Sendai (Kyushu Electric)

Nuclear reactor status

- Accident with nuclear fuel damage (4)
- Accident without nuclear fuel damage (2)
- Safe shutdown (8)
- Not affected by earthquakes (40)
- Under construction (2)
- Construction planned (10)

2 Energy mix

Oil | Coal | Gas | Nuclear | Renewables

2007: 136 | 261 | 282 | 264 | 88

2030 (estimate): 113 | 136 | 537 | 214 | 21

Power generation, TWh

▶ firms say they will stick to construction schedules to meet the nuclear capacity goals, the Japanese government has ordered a review of its energy policy. "They obviously have a predicament in that they have put a lot of eggs in the nuclear-power basket," says Ed Lyman, a nuclear expert with the Union of Concerned Scientists in Washington DC. "It's hard to imagine that Fukushima isn't going to disrupt those plans."

The most immediate challenge is to meet the nation's electricity needs. The rolling blackouts introduced after the crisis began have stopped, thanks to reduced demand and energy conservation measures. But supplying enough power is difficult because Japan maintains two essentially incompatible electricity grids, a decision dating back more than a century. These grids carry alternating current at different frequencies, preventing large-scale diversion of electricity from power plants outside the affected region. There are no plans to unify the systems.

In the short term, most experts agree that the crisis will force Japan to rely much more heavily on its natural-gas power stations, and potentially also its coal-fired plants. But increasing fossil-fuel use probably won't stop Japan from meeting its short-term carbon-emissions commitment under the Kyoto Protocol — a reduction of 6% below 1990 levels by 2012 — says Lisa Zelljadt, a senior analyst with the energy consultancy Thompson Reuters Point Carbon in Washington DC.

Japan's emissions have increased significantly since 1990, so the country has been working to offset much of its emissions by paying for reductions in other countries. "The only way that they were going to make their Kyoto commitments was by buying carbon credits," says Robert Stavins, an economist at Harvard University in Cambridge, Massachusetts. Japan also moved a little closer towards its climate targets during the global economic crisis, when plunging economic output reduced emissions by 7% from 2007 to 2008. The earthquake, tsunami and nuclear crisis will also help curb carbon output.

All these factors should put Japan in a position to meet its Kyoto commitments, says Zelljadt. "The nuclear disaster will thus affect more how Japan stands towards a longer-term target in the international negotiations," she says.

In a move towards not just carbon-free energy but also energy self-sufficiency, Japan had outlined ambitious plans to develop an advanced nuclear-energy programme that would not only recycle spent nuclear fuel but also deploy fast-breeder reactors that produce more fuel than they burn (see *Nature* **464,** 661; 2010).

These efforts are a cornerstone of Japan's climate strategy, but they were under pressure even before the earthquake. Japan's Rokkasho reprocessing plant has suffered from numerous problems and will be some 15 years behind schedule if it opens as currently planned in 2012. The plant stores thousands of tonnes of spent nuclear fuel, and had to rely on backup diesel generators to cool the fuel after the plant lost power following last week's big aftershock.

Meanwhile, a prototype fast-breeder reactor — dubbed Monju — has been plagued with problems (see *Nature* **434,** 6; 2005). Fifteen years after a sodium coolant leak and a related fire led to its closure, Monju was briefly started up in 2010 only to be shut down once again after a non-nuclear accident.

"It's a pretty grim story anywhere you look in Japan," says Lyman. "You have to wonder how many nuclear failures they can put up with before throwing in the towel." ∎

Brown seaweed, usually harvested off the coast of Japan (above), is absorbing radioactive iodine.

RADIOECOLOGY

Radiation release will hit marine life

Researchers call for extensive surveys to gauge ecological effects of Fukushima.

BY QUIRIN SCHIERMEIER

As radioisotopes pour into the sea from the crippled Fukushima Daiichi nuclear plant, one reassuring message has been heard over and over again: the Pacific Ocean is a big place.

That the isotopes will be vastly diluted is not in question. Nevertheless, scientists are calling for a marine survey to begin as soon as possible to assess any damage to ecosystems in the area around Fukushima. Although the contamination is unlikely to cause immediate harm to marine organisms, long-lived isotopes are expected to accumulate in the food chain and may cause problems such as increased mortality in fish and marine-mammal populations.

"Just because you can measure it, doesn't mean it's dangerous," says Ken Buesseler, a marine geochemist at the Woods Hole Oceanographic Institution in Massachusetts. "Even so, this is the biggest man-made release ever of radioactive material into the oceans. We haven't yet seen enough data to assess what's going on, so anything that can be done in terms

of further monitoring would be very welcome."

The past two weeks have seen extremely high concentrations of radioactive iodine-131 (with a half life of 8 days) and caesium-137 (which has a half life of 30 years) in samples of sea water collected near the Fukushima reactors, and even as far as 30 kilometres offshore. By

late March, levels were tens of thousands of times higher than before the accident (see 'Radio-isotope contamination'). Many other radio-isotopes, both long- and short-lived, are also likely to have been released.

But the total amount of radioactivity that has entered the ocean is unknown, and discharges — both accidental and deliberate — are continuing and may even be substantial if any further problems occur at the Fukushima plant (see page 146).

Despite these uncertainties, scientists at the National Institute of Radiological Sciences (NIRS) in Chiba, Japan, are designing studies aimed at monitoring the accumulation of radionuclides in the muscles, organs, eggs and bones of marine organisms. They also plan to model the long-term behaviour of radioisotopes in the marine environment, and the overall radiation doses to which marine organisms will be exposed. "We need to obtain specific concentrations of caesium and iodine isotopes to assess their effects in different marine organisms," says Tatsuo Aono, an expert in marine radioecology at the NIRS.

A team led by Dominique Boust, director of the French Institute of Radioprotection and Nuclear Safety (IRSN) in Cherbourg, is now predicting the level of contamination in marine organisms and sediments using estimates of the quantity of radioisotopes released from Fukushima, and the ratios of those isotopes calculated from available sea-water measurements.

The team calculates that about 50 radio-isotopes contribute to an overall concentration of roughly 10,000 becquerels per litre in the sea water within 300 metres of Fukushima. Before the accident, caesium-137 concentrations there were about 0.003 becquerels per litre, and iodine-131 was not detectable. On the basis of these figures, the IRSN researchers suggest that sediments in the region could now contain 10,000–10 million becquerels ▶

RADIOISOTOPE CONTAMINATION

Levels of caesium-137 and iodine-131 in the sea close to the damaged Fukushima reactors have shattered legal limits (40 becquerels per litre for iodine-131 and 90 becquerels per litre for caesium-137).
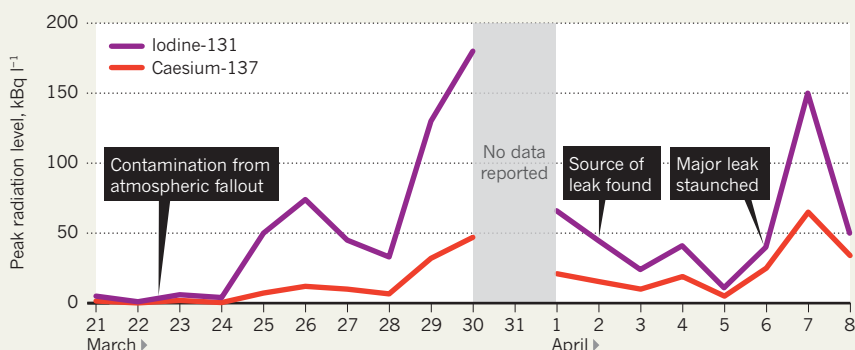
▶ per kilogram; fish could carry 10,000–100,000 becquerels per kilogram; and algae, some of which are particularly susceptible to iodine uptake, could contain up to 100 million becquerels per kilogram. Japan has legal limits of radioactivity in fish for human consumption of 500 becquerels per kilogram for caesium-137, and 2,000 becquerels per kilogram for iodine-131.

"Doses will decrease very quickly with time and distance from the facility, if no further leaks occur, but there could remain a persistent low-dose component in the local marine environment for many years," says Thomas Hinton, deputy director of the IRSN's Laboratory of Radioecology, Ecotoxicology and Environmental Modelling in Cadarache, France. "The impacts are best addressed through an international long-term assessment."

Ward Whicker, an environmental and radiological health expert at Colorado State University in Fort Collins, agrees that a survey would be worthwhile. "It would require a great deal of sampling effort, near the discharge point as well as at locations farther away," he says. "Concentrations of radionuclides in water, sediments, plankton, molluscs, crustaceans, seaweed and fish would need to be measured, and the health of the ecosystem monitored."
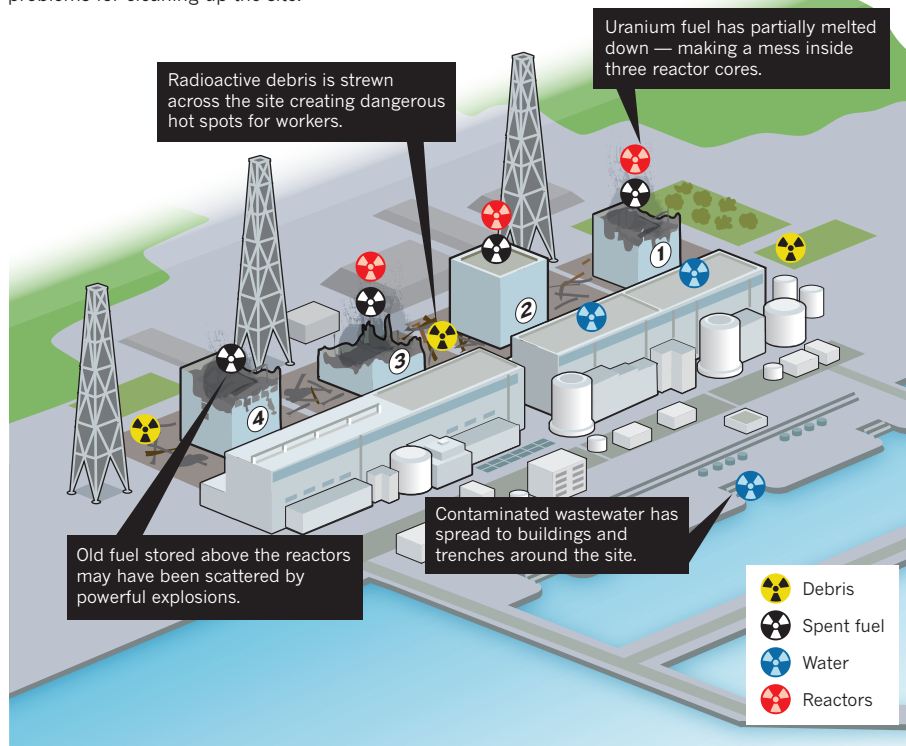
Although radioisotope concentrations in fish, shellfish and seaweed could exceed limits for human consumption for weeks, Whicker thinks that it is unlikely that scientists would be able to detect any genetic effects on marine life. Any affected creatures would probably disperse into the Pacific, or die more quickly, he says. Moreover, teasing out radiological effects from other stresses, such as conventional water pollution and the damage caused by the tsunami, would be extremely difficult.

An alternative approach could be to focus on a suitable proxy species. "In my opinion, brown seaweed should be the number one research priority," says Bruno Fievet, a radioecologist at the IRNS in Cherbourg. The brown seaweed *Laminaria digitata*, ubiquitous in the coastal Pacific waters off Japan, absorbs iodine to help defend itself against environmental stresses such as pollution. It can have iodine concentrations some 10,000 times greater than the surrounding sea water. "This species is the world champion in iodine uptake, and it would be a good indicator of the radioactive labelling of other marine organisms," says Fievet.

But sampling may be hampered by the danger that remains at the Fukushima plant. "Any survey would be welcome," says Ulf Riebesell, a biological oceanographer at the Leibniz Institute of Marine Sciences in Kiel, Germany. "But I certainly wouldn't ask my students to do field work off Japan amid this ongoing crisis." ∎

FUKUSHIMA'S RADIOACTIVE WRECKAGE

Radioactive contamination takes many forms at Japan's stricken nuclear reactors. Each source creates its own problems for cleaning up the site.

Radioactive debris is strewn across the site creating dangerous hot spots for workers.

Uranium fuel has partially melted down — making a mess inside three reactor cores.

Old fuel stored above the reactors may have been scattered by powerful explosions.

Contaminated wastewater has spread to buildings and trenches around the site.

Debris
Spent fuel
Water
Reactors

NUCLEAR ACCIDENT

Fukushima set for epic clean-up

Latest data suggest a Chernobyl–like effort will be needed.

BY GEOFF BRUMFIEL

As the immediate threat from Fukushima Daiichi's damaged nuclear reactors recedes, engineers and scientists are facing up to a clean-up process that could last for many decades, or even a century.

Experts on previous nuclear accidents say that the sheer quantity of nuclear material that needs to be removed from the site, together with the extent of the damage, makes Fukushima a unique challenge. The plant's damaged reactors are home to just under 1,000 tonnes of nuclear fuel and thousands of tonnes of radioactive water (see graphic).

Last week, the Toshiba Corporation floated a rough proposal to clean up the site in a decade. But veterans of clean-up operations at sites such as Three Mile Island in Pennsylvania say that it will probably take much longer. The removal of the radioactive material will require a carefully planned and technologically sophisticated programme, made all the more challenging by the devastation left after partial core meltdowns and explosions.

No clean-up can begin until the reactors are stabilized. Radiation around the plant is beginning to wane, but the threat of further releases has not yet passed. On 7 and 11 April, severe aftershocks struck nearby, raising fears that the three crippled reactors could be damaged further. The Tokyo Electric Power Company (TEPCO), which manages the plant, says that no additional damage has been detected.

A 26 March report from the US Nuclear Regulatory Commission (NRC), leaked to *The New York Times*, says that massive explosions at the plant in March scattered fuel from the reactors' spent-fuel pools around the site. NRC officials also believe that a portion of the uranium fuel inside the unit 2 reactor may have escaped its stainless steel containment

vessel and fallen onto the concrete floor below, although the Japanese government has yet to confirm this. In addition, the document indicates that water is not circulating properly through the cores of the damaged reactors, so it will be necessary to continue to flood them, says Richard Lahey, an emeritus professor of nuclear engineering at Rensselaer Polytechnic Institute in Troy, New York.

This strategy creates its own problems. The reactors' cooling systems are normally a closed circuit. Flooding the cores means that water contaminated with radioisotopes will continue to spill out into the environment. TEPCO has already reported highly radioactive water in buildings and trenches around the site.

Dealing with the water will be a pressing priority for any clean-up operation, according to Jack DeVine, an independent nuclear consultant who spent six years dismantling the unit 2 reactor at Three Mile Island after it partially melted down in 1979. The accident left thousands of tonnes of water laced with radioactive caesium-137 swilling around in the reactor's basement. Over the course of months, the US team built a system that could suck the water out and pass it through radiation-resistant zeolite filters. The zeolite removed the caesium and other radioisotopes, leaving almost pristine water, which was eventually evaporated at a facility on the site. A similar system could work at Fukushima, says DeVine, although the constant leakage from the damaged cores means that any clean-up is a race against time. More than 10,000 tonnes of low-level radioactive water has already had to be dumped from storage tanks into the Pacific Ocean to make way for more-radioactive cooling water (see page 145).

Cleaning up the reactors themselves presents an even greater challenge. Debris and high radiation levels are making it impossible to conduct proper surveys of the damage. In the near term, robots will need to explore the reactor buildings and map the radiation inside, says Red Whittaker, a robotics expert at Carnegie Mellon University in Pittsburgh, Pennsylvania, who has developed systems for other nuclear accidents.

It could be years before anyone can look inside the cores themselves. At Three Mile Island, engineers had to wait three years before radiation levels had fallen sufficiently to allow them to lower a camera through a control rod drive shaft into the heart of the reactor. At Fukushima, it could take longer still. The boiling water reactor (BWR) design used there is sealed with a solid stainless-steel cap that can only be removed by a heavy, fuel-loading crane located above the reactor. Explosions at three of the units with fuel in their cores mean that "their cranes are

clearly toast", says DeVine. Plant operators will have to find another way in.

The reactor's design also presents other problems. "The BWR is just a rat's nest of tight spaces, pipes and valves," says DeVine. To effectively remove the fuel, DeVine thinks that one or more new buildings with dedicated cranes must be built around each reactor. "It's not something that will be up in a month or two," he says. Whittaker adds that robots and humans will need to share the work in a methodical, coordinated way. "The nature of these operations is that they are patient and persistent," he says.

Indeed, the effort required seems likely to be more akin to the clean-up strategy at Chernobyl in the Ukraine than that of Three Mile Island. Engineers at Chernobyl are beginning to lay the foundations for a massive €1-billion (US$1.4-billion) enclosure, complete with automated

"Bottling it up and leaving it seems to me to be a really bad choice."

cranes, that will eventually lift apart the sarcophagus of steel and concrete hastily thrown up around the stricken unit 4 reactor in the months after it exploded in 1986. The new building, which is intended to last a century, was agreed in principle in 2001, but will not be completed until at least 2015. Clean-up of the site is scheduled to last until 2065 — almost 80 years after the accident.

TEPCO almost certainly cannot afford a clean-up on this scale. "I think that, ultimately, the government is going to have to pay for it," says Robert Alvarez at the Institute for Policy Studies in Washington DC, who oversaw clean-up of former US nuclear weapons plants during the administration of President Bill Clinton. The government already seems to be mulling whether to take over the utility, shares in which have plunged since the accident.

Given the complexity of the task ahead, some think it may be better to abandon Fukushima entirely — at least for the time being. "My bet would be: you seal it and wait a hundred years," says Alan Johnson, a retired reactor physicist who was head of Britain's Sellafield nuclear processing site in the late 1980s.

Sellafield, once known as Windscale, was in 1957 the site of the United Kingdom's worst nuclear accident, when a reactor's graphite core caught fire. Final decommissioning of the reactor is still at least 20 years away, but the hiatus has allowed radioactive materials to decay and given engineers time to develop the best clean-up strategies possible. "What's the rush in doing it quicker?" asks Johnson.

But natural disasters are rare in England. Given the threat of major earthquakes, tsunamis and typhoons that could strike Japan in the decades to come, DeVine has his doubts about applying the same strategy. "Bottling it up and leaving it seems to me to be a really bad choice," he says. ∎

PATENT REFORM

US legislation aims to simplify rules for inventors

Bill aims to reduce costly lawsuits by changing how rights to innovations are awarded.

BY EUGENIE SAMUEL REICH

The US Congress is poised to enact the most sweeping overhaul of the country's patent legislation for almost 30 years. Advocates say that a law currently under consideration will make it easier for scientists to commercialize patents without getting caught up in costly litigation. "After more than six years, we have suddenly made rather important progress," says John Vaughn, executive vice-president of the Association of American Universities in Washington DC, who has spearheaded the organization's repeated efforts to bring patent reform to a vote.

The America Invents Act was introduced into the House of Representatives on 30 March and is scheduled to be taken up in committee this week; a similar bill enjoyed swift passage through the Senate on 8 March. The act is the culmination of a long campaign by universities and companies to change the US patent system to a first-to-file arrangement, in line with most of the rest of the world. At the moment, patents in the country are awarded on a first-to-invent basis, so that if two inventors file similar patent applications at around the same time they go into 'interference', in which a specialized division of the US Patent and Trademark Office attempts to decide who came up with the invention first. It costs an average of US$400,000–500,000 to fight an interference case — more than most academic start-ups can afford.

"Interferences are extremely expensive and rather uncertain," says Lita Nelsen, director of technology licensing at the Massachusetts Institute of Technology (MIT) in Cambridge. "I'd rather live with the certainty and lose quickly." She says that the new rule won't be a big adjustment for MIT, which routinely seeks patent rights outside the United States, in countries that are already following the first-to-file system.

Defenders of first-to-invent have argued that it benefits individual inventors who don't manage to file their patents quickly. But in testimony to Congress on 30 March, David Kappos, director of the US Patent and Trademark Office, said that only one of the three million

inventions handled by the office over the past seven years had involved an individual inventor who was second to file but was able to demonstrate that he was first to invent, and was therefore awarded the patent. Others who prevailed through an interference were companies or groups of inventors.

"This has been a red herring used by people who have an interest in keeping the patent system as complicated as possible; that is, patent lawyers," says Josh Lerner, who studies innovation policy at Harvard Business School in Boston, Massachusetts.

But Gail Naughton, a dean of business administration at San Diego State University in California who served on a National Academies panel that recommended the shift to first-to-file in 2004, is now concerned that the change could trigger a surge of filings as inventors rush to cross the line first (see 'Patents pile up'). "If substantial funding to support many new hires at the patent office is not made available, the change could result in an even greater backlog," she says, adding that, under the new system, it would be important for universities to train academics in how to identify and obtain the data vital to describing an invention in a patent, so that they can file quickly but not prematurely.

Rob Merges, who studies patent law at the University of California, Berkeley, says that the new law could hurt academics because it weakens protections for inventors who start commercializing innovations before filing

a patent. "I'm not thrilled with that," says Merges, who adds that scientists often need to send manufactured samples to contacts in industry to advance their work before they can formally seek a patent.

Universities are also uneasy with a provision in the House bill that would establish 'prior user rights'. These would protect parties who developed an innovation first, but failed to patent it themselves, against claims by other parties who later patented the same invention. "Prior user rights are essential to prevent foreign and non-practising entities from filing patents on American inventions," says Mark Chandler, general counsel of Cisco Systems in San Jose, California.

> *"Interferences are extremely expensive and rather uncertain. I'd rather live with the certainty and lose quickly."*

Vaughn, however, says that the provision could work against academics, by allowing companies who have independently developed university-patented technologies to use them without penalty. Furthermore, academics would not benefit from the rights, he says, because they tend to make money by licensing patented technology rather than by making products themselves. The House bill does contain an exemption that would prevent anyone from claiming prior user rights to inventions that are wholly university-funded. But Vaughn says that such precautions don't go far enough to protect universities' rights, because some university inventions are developed using private capital.

One provision of the law that has not been heavily disputed would introduce a post-grant review process to let outsiders challenge newly awarded patents without costly litigation. This would further bring the United States into line with the European patent office, which already has an opposition procedure to enable people to challenge new patents. "We think post-grant review is a quality check and we support that," says Gary Griswold of the Coalition for 21st Century Patent Reform, a group of nearly 50 US and multinational corporations from various industries, including several pharmaceutical companies. Lerner agrees. "The real secret is having people nip bad patents in the bud," he says. ∎

PATENTS PILE UP

The number of US patent applications each year has doubled in a decade, causing a backlog.
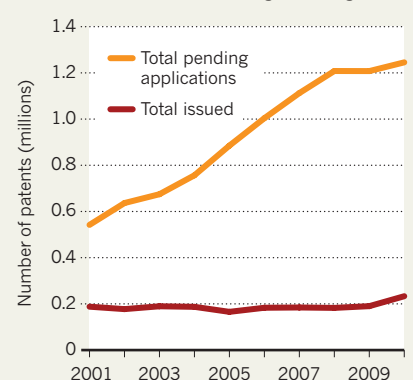
SOURCE: US PATENT AND TRADEMARK OFFICE

Online archives can host 3D scans of fossils, such as the arachnid *Eophrynus prestvicii* (above).

R. GARWOOD

PALAEONTOLOGY

Fossil data enter the web period

Palaeontologists call for more sharing of raw information.

BY EWEN CALLAWAY

When Ross Mounce tried to use measurements of a fossilized feathered theropod dinosaur in his research, he hit a big stumbling block. A data table in the paper describing the dinosaur contained a formatting error that prevented Mounce from working out the creature's relationship to other dinosaurs, and the paper's senior author ignored repeated e-mail requests for the original file.

Mounce, who is studying for a PhD in evolution at the University of Bath, UK, eventually got the file after the journal's editor intervened. Now Mounce is leading a campaign to avoid such situations by making it standard practice for palaeontologists to upload the raw data behind their papers into online repositories — common practice in other disciplines but rare in palaeontology. The call has drawn a mixture of support and dismay, but a sea change may already be under way: several palaeontology journals have recently rolled out digital-archiving policies that align with Mounce's goal.

Palaeontologists calling for better digital

NATURE.COM
For the latest palaeontology news from *Nature* see:
go.nature.com/w8murr

record-keeping and data sharing say that the move will not only spare scientists the trouble of tracking down information about fossils, but will also underpin future studies using the data in ways that are not possible today. "I think that is far and away the biggest plus," says Peter Wagner, curator of palaeozoic molluscs at the US National Museum of Natural History (NMNH) in Washington DC, who backs the campaign.

Public repositories already aid data sharing in other fields. Genbank (www.ncbi.nlm.nih.gov/genbank), a database run by the US National Center for Biotechnology Information in Bethesda, Maryland, makes more than 100 million genetic sequences freely available, and most journals demand that scientists publishing in their pages add their data to it.

Similar databases do exist for palaeontologists. Morphobank (morphobank.org), for example, logs detailed fossil images, including three-dimensional computed tomography scans. Another repository, Treebase (treebase.org), collects the phylogenetic trees that show evolutionary relationships between species. But journals do not compel researchers to add their data to these systems, and Mounce — along with almost 150 signatories to an open letter (supportpalaeodataarchiving.co.uk) — thinks that they should.

Some scientists are concerned that disclosing key data such as detailed geographical information about fossil discoveries would feed the illicit fossil trade. "The Burgess Shale is a case in point. It's a well known locality and it's pillaged to death," says Jonathan Antcliffe, a palaeobiologist at the University of Bristol, UK, referring to a rich trove of Cambrian fossils in the Canadian Rocky Mountains. Some countries, including the United States, make it illegal to publicly disclose fossil sites on public lands.

Antcliffe is also concerned that mandatory data archiving could discourage scientists from publishing progress reports on long-term projects, fearing that rivals will use the raw data to scoop them. He adds that graduate students, who might take more time than experienced palaeontologists to turn research into papers, would be especially vulnerable.

Tensions between scientists who discover new fossils and those who analyse and synthesize their finds are not new, says Mike Benton, a vertebrate palaeontologist at the University of Bristol. For example, Jack Sepkoski of the University of Chicago, Illinois, who in the 1970s and 1980s studied mass extinctions in the global fossil record, faced criticisms for repurposing other scientists' field work. But, says Benton, "if you wanted to keep it secret, you shouldn't have published it".

Whether palaeontologists are ready for mandatory digital archiving or not, the field seems to be going in that direction. Propelled in part by data-sharing edicts from funding agencies such as the US National Science Foundation, the *Journal of Vertebrate Paleontology* announced in January that it would require authors to post raw data files on its website (A. Berta and P. M. Barrett *J. Vert. Paleontol.* **31,** 1; 2011). It is also considering mandating storage in public repositories such as Morphobank. Meanwhile, the Paleontological Society in Boulder, Colorado, which publishes *Paleobiology* and the *Journal of Paleontology*, last month decided to archive data from its papers using a repository called Dryad (datadryad.org). "My only concern is that archiving so far is an unfunded mandate," says Philip Gingerich, the society's president. "Archiving could easily consume an entire research budget."

Brian Huber, curator of planktic foraminifera at the NMNH and co-editor of the *Journal of Paleontology*, says that he too was wary of the costs of digital archiving, but has come around to the idea. "This is the way of the future, and the society decided we've got to lead on this instead of being too conservative." ∎

CORRECTION

The News story 'Bulgarian funding agency accused of poor practice' (*Nature* **472,** 19; 2011) wrongly stated that 250 million leva is equivalent to US$181,000. It should have been US$181 million.
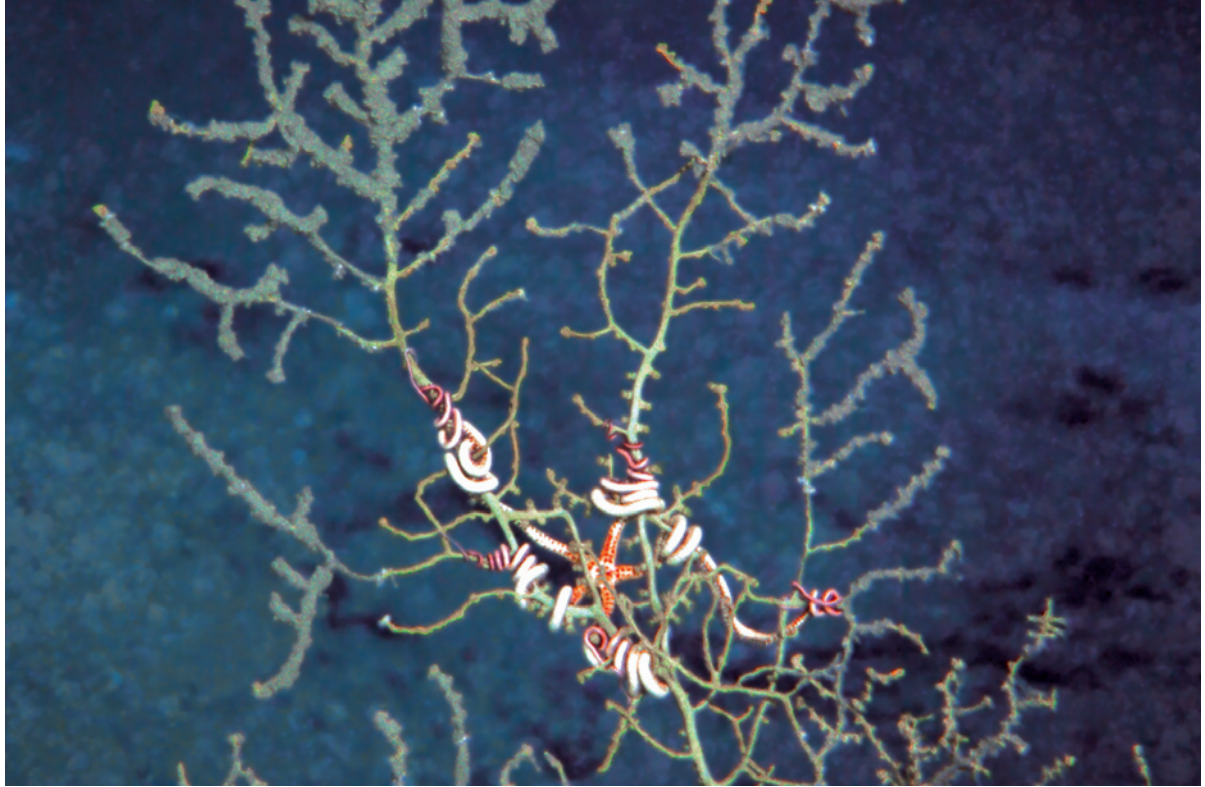
Peculiar dark material covers much of this coral from the Gulf of Mexico. A sickly brittle star clings to the coral, little of which is still alive.

DEEP WOUNDS

The Gulf of Mexico oil spill set records for its size and depth. A year on, the biggest impacts seem to be where they are hardest to spot.

BY MARK SCHROPE

L ate last year, oceanographers prowling the sea floor of the Gulf of Mexico came upon what looked like a crime scene. Cameras on a remotely operated vehicle revealed corpses of deepwater coral covered in brown goo. As the researchers watching from above saw one grim scene after another, "the whole place got silent, everything totally stopped", says Tim Shank an oceanographer from the Woods Hole Oceanographic Institution in Massachusetts, and a member of the survey team.

The field of coral was just 11 kilometres from the Deepwater Horizon well head, which earlier in the year had spewed out more than 4 million barrels of oil and a similar amount of methane — the largest ever accidental release in the ocean. The spill was unique in other ways, too. Located beyond the continental shelf and some 1,400 metres below the surface, it happened in deeper water than any other major spill in history.

Those factors make it much harder for researchers to discern what happened to the oil and how it affected wildlife. Assessing the impacts of this spill in deep water is "probably one of the most challenging things ever", says Steve Murawski, chief scientist for the Fisheries Service at the National Oceanic and Atmospheric Administration (NOAA) during the spill and now a fishery biologist at the University of South Florida in St Petersburg. And that is "not only because of the physical environment but also because of the breadth of the potential impacts", he says.

But tracking the oil and its impacts remains an essential task. It will help to determine how much needs to be paid in restoration costs by BP — the company in charge of the well — and possibly other companies deemed partially responsible. And as drilling by the United States and other countries expands into deeper waters, lessons from last year's spill could help in responding to any future ones.

At first, scientists and the public were most concerned about how the disaster would harm coastlines and near-shore waters. Although those areas sustained some damage, they did not come to as much harm as many had feared. Instead, as the anniversary of the spill approaches, signs of significant damage are showing up farther from shore and in deeper water. It was a stroke of bad luck that the well happened to be located in the most species-rich part of the deep gulf.

Thomas Shirley, a veteran of spill research at the Harte Research Institute at Texas A&M University in Corpus Christi, says that his view of the spill is evolving. "I'm beginning to think the deep sea is where we'll see most of the effects," he says.

BEYOND THE EDGE

True to its name, the Deepwater Horizon oil rig was stationed beyond the lip of the continental shelf, where the sea floor rapidly falls away towards the lower reaches of the Gulf of Mexico. On 20 April 2010, a catastrophic blowout caused an explosion on the oil rig that claimed 11 lives and sent the rig to the sea floor. Government and independent estimates calculate that the broken well sprayed 4.1 million barrels of oil — and perhaps as much as 363,000 tonnes of natural gas — into the deep waters of the Gulf of Mexico.

Operations to collect and burn the oil took care of only about one-quarter of the liquid that came out of the well, according to a controversial government report[1]. The rest dissolved into the sea, dispersed into small drops, evaporated into the atmosphere or initially formed visible surface slicks and tar balls (see 'What happened to the oil?').

Some of it apparently went through an unusual transformation. Scientists who were out on boats in the spill zone during the early weeks after the disaster saw unusual strings of viscous material, which

they dubbed 'sea snot'. The material looked like oil mixed with phytoplankton and other organic matter — and resembled a very thick batch of egg-drop soup.

Vernon Asper, a geochemist at the University of Southern Mississippi near Diamondhead, had never seen anything like it. "What is becoming of all that," he asked in mid-May while looking into water just a few kilometres from the blowout site. "Where is it going?" Asper and others found that the substance extended well below the surface in places.

"This is very strange material," says Ed Overton, a marine chemist and spill expert at Louisiana State University in Baton Rouge. But although many researchers reported seeing it, few were equipped to collect it. "We had one hell of a time getting good samples," says Overton. Although his team is still analysing samples obtained from other researchers, the group has already made some intriguing findings that could explain the bizarre appearance.

In shallow spills, oil tends to rise quickly to the surface, where it weathers, dissolves and evaporates in chemically predictable patterns. However, the largest drops of oil from the Deepwater Horizon well head took at least four hours to reach the surface, and smaller droplets rose much more slowly. During that long voyage, smaller droplets could have lost some of the lighter hydrocarbons that help to keep the various oil compounds from separating, suggests Overton.

"That changes the properties of the oil so it goes from a nice little round droplet, I think, into these strange-looking filamentous globs floating up the water column," he says. "Unfortunately we didn't get enough oil to really harden this theory."

Still, the issue of how the oil transformed is a crucial one for researchers to address. The processes involved can affect the oil's toxicity and how long it is likely to stick around.

If Overton is right, the stringy masses that researchers found were a reminder that Deepwater Horizon was a different kind of spill. Soon after the disaster, Asper and his colleagues started seeing indications that some oil never made it to the surface — instead, it formed diffuse plumes more than 1,000 metres below the surface[2]. Reports of those finds were initially greeted with scepticism, in part because BP and most government officials expected that all the oil would float. But Asper's group and other researchers eventually confirmed that an unknown fraction of the oil had drifted away from the well head in deep-water plumes[3], the effects of which are still not clear.

Since then, controversy has also erupted over how much oil settled on the ocean floor and in what form. Samantha Joye, a geochemist at the University of Georgia in Athens is part of a team that found a layer of brown lumpy material on the sea floor that she says looks like dirty cauliflower and can be seen at sites as far as 130 kilometres from the well head.

Andreas Teske, a microbiologist at the University of North Carolina, Chapel Hill, collaborates with Joye and suggests that the lumpy layer came from oil that had once been at the surface, perhaps beginning as the stringy sea snot. As microbes consumed the oil, it would have lost buoyancy and sunk, he says. Kai Ziervogel, a biogeochemist also at the University of North Carolina, tested this hypothesis in the laboratory by incubating seawater samples with surface oil from the spill. He saw masses of bacteria and oil forming, some of which sank and looked much like the lumpy layer of material in the sediments.

Teske's team analysed samples from the lumpy layer and found that they contained oil with an unusually high percentage of the heavier compounds that would have been hard for microbes to eat, which fits with the idea that bacteria had consumed the lighter compounds. The group has also dated the lumpy layer — using the radioactive decay of thorium-234 as a clock — and found that it formed during and shortly after the spill.

Joye has called the cauliflower layer a graveyard because she and her colleagues found in it countless dead worms and other common sediment dwellers, as well as the remains of jellyfish and other animals from above. And near to the well head the layer shows little microbial activity, suggesting that it will not break down quickly.

The group has not yet published these findings and has just started to describe them at scientific conferences. But a number of researchers have questioned the existence of a widespread lumpy layer containing oil. Arden Ahnell, gulf-coast restoration science manager for BP, says that researchers working with the company have confirmed low concentrations of oil in some of the areas where Joye and her colleagues collected sediment, but found no evidence of a pervasive layer of unusual material.

Joye has grown used to people challenging her work, having been part of the group that first reported the existence of a deep oil plume. In her laboratory, she points out the difference between normal grey sediments collected from deeper layers near the spill site and the lumpy brown sediments that, she says, formed after the spill. "These are not normal-looking sediments. They are the most putrified, ugly, nasty-looking things you could imagine."

Her arguments get some support from Amanda Demopoulos, a benthic ecologist with the US Geological Survey in Gainesville, Florida. Demopoulos, too, found an unusual lumpy brown deposit at one of the sites later visited by Joye and her colleagues. In Demopoulos's samples, "there were some snails still living but the remaining animals I found were not moving around, and that's strange", she says.
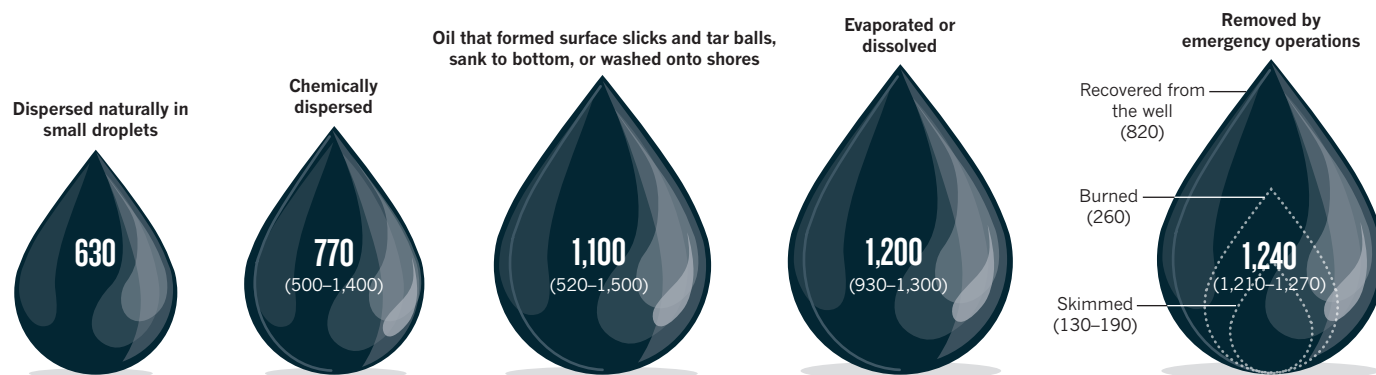
BUILDING A CASE

Some of these data may end up in court, as part of a process called the Natural Resource Damage Assessment (NRDA). The US government will use studies conducted through the NRDA to document the effects of the spill and then determine what damages should be paid by the responsible companies. This will include reimbursing the government for the cost of the NRDA.
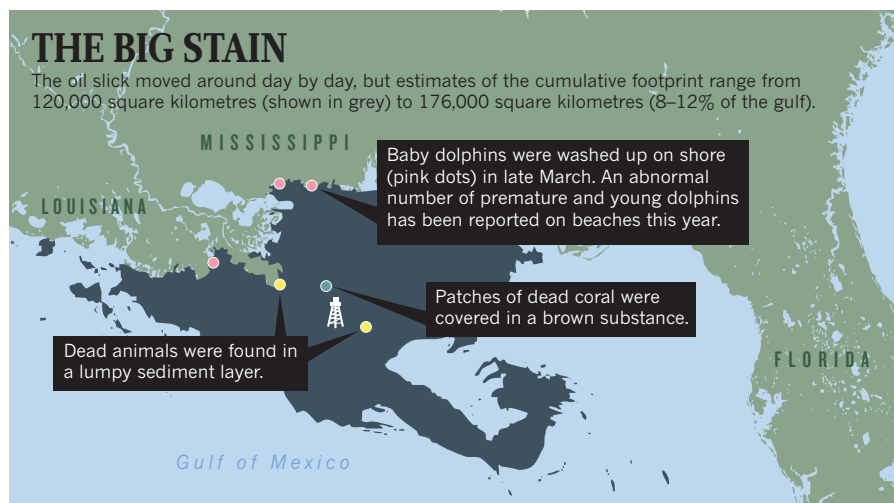
Because of the legal nature of the work, many of the academic

WHAT HAPPENED TO THE OIL?

During the Deepwater Horizon crisis last year, the US government estimated where the 4.9 million barrels of oil went so it could plan response efforts. In November, it issued revised numbers, as well as ranges for some categories (shown in parentheses). All numbers are in thousands of barrels.

SOURCE: REF. 1

Dispersed naturally in small droplets
630

Chemically dispersed
770
(500–1,400)

Oil that formed surface slicks and tar balls, sank to bottom, or washed onto shores
1,100
(520–1,500)

Evaporated or dissolved
1,200
(930–1,300)

Removed by emergency operations
1,240
(1,210–1,270)

Recovered from the well (820)

Burned (260)

Skimmed (130–190)

THE BIG STAIN

The oil slick moved around day by day, but estimates of the cumulative footprint range from 120,000 square kilometres (shown in grey) to 176,000 square kilometres (8–12% of the gulf).

MISSISSIPPI

LOUISIANA

Baby dolphins were washed up on shore (pink dots) in late March. An abnormal number of premature and young dolphins has been reported on beaches this year.

FLORIDA

Patches of dead coral were covered in a brown substance.

Dead animals were found in a lumpy sediment layer.

Gulf of Mexico

researchers who are participating in it find the process foreign. Some signed non-disclosure forms that prevent them from publicly discussing their results without approval. And the researchers must maintain strict procedures for handling samples.

Erik Cordes, a deep-sea ecologist from Temple University in Philadelphia, Pennsylvania, was one of the leaders of the coral study and he expressed frustration early on with the NRDA process. Cordes had wanted to conduct a well-accepted analysis of stress on the seafloor animals but was told by NRDA legal advisers that the work might not be accepted because the techniques had not been proven in court. "The idea that there was a higher standard than peer review and scientific consensus is very difficult to accept," says Cordes.

The NRDA is a major source of funding for research into the effects of the oil spill. Scientists should eventually gain access to around $400 million in research money from BP, but for now that fund has released only limited amounts because of bureaucratic and political problems[4].

A significant amount of BP's money will go towards studies of how the spill damaged gulf ecosystems. Its location 66 kilometres offshore might have helped to limit damage to the estuaries and coastal waters, but it was still located in a region of relatively rich biodiversity, says Shirley. He and his colleagues have amassed a database of 15,419 species living within the gulf. According to the database, some 1,728 species inhabit the region surrounding Deepwater Horizon at depths of between 1,000 and 3,000 metres, where the well is located. "Ironically, this is the most speciose area of the gulf for this depth range — BP could not have selected a worse area to have a spill, at least from the point of species richness," says Shirley.

Researchers are finding signs that the damage extends throughout the water column, from the sea floor to the surface (see 'The big stain'). During the expedition on which Joye and others made the sediment discoveries, other researchers were concerned by jellyfish and other creatures they pulled up in nets from lower depths as far as 150 kilometres from the well head. "The gelatinous animals are usually pinkish or translucent, but an awful lot of them were much darker brown or even black," says Joseph Montoya, chief scientist on the cruise, and a biological oceanographer at the Georgia Institute of Technology in Atlanta.

In the spill zone soon after the disaster began, Asper and his colleagues often pulled up their equipment and found it decorated with the dead bodies of invertebrates known as pyrosomes, or fire salps. Pyrosomes probably have an important role in the food web near the surface, says Shirley, because they consume small plankton.

Aside from those more obvious casualties, researchers have been hard pressed to find mass deaths that are clearly linked to the spill. This

year, the bodies of 151 bottlenose dolphins have washed ashore in the northern gulf, but mass dolphin strandings have happened there before. The number of animals with clear evidence of oil contamination is much smaller. Only 6 dolphins and 18 sea turtles have been found dead with visible signs of oil, according to NOAA.

Researchers caution that the oil spill probably took a much larger toll on whales, dolphins and turtles than has been observed. Blair Witherington, a sea-turtle expert with the Florida Fish and Wildlife Conservation Commission in Melbourne Beach, Florida, is particularly concerned about juvenile Kemp's ridleys (*Lepidochelys kempii*), the rarest species of sea turtle. Young turtles are the most vulnerable, he says, and are generally found well offshore. His team found hundreds that had visible signs of oil.

For dolphins and other cetaceans, the gap between known deaths and actual ones may be vast. Judging from past studies of death rates, a group of cetacean specialists concluded last month that "the true death toll could be 50 times the number of carcasses recovered"[5].

THE MOST VULNERABLE

Researchers are especially concerned about the youngest creatures because the spill came at a time when many animals were spawning. Larvae, for example, are known to be particularly susceptible to the toxic chemicals in the oil. And the problems may have been compounded by the 3 million litres of dispersants that were released to break up the oil. The dispersants themselves can be toxic, and they make oil droplets smaller and therefore more likely to affect even the smallest creatures.

Ed Stellwag, a developmental biologist at East Carolina University in Greenville, North Carolina, is exploring what mixtures of spill oil and dispersant do to the embryos of zebrafish, the aquatic equivalent of lab rats. He found that even at the modest concentrations that many animals would have encountered during the spill, the mixture caused fatal heart and other defects in all of the embryos tested. Stellwag says that assessing how embryonic damage plays out offshore is likely to be difficult if not impossible because larger animals eat most of the larvae in surface waters. Damaged larvae, he fears, are more likely to end up as a meal than in a researcher's survey net.

> **The Deepwater Horizon site is now strangely quiet.**

Larva surveys are already under way as part of the NRDA process. But given the difficulty of the work and the controls on releasing information, it could be years before researchers can offer an assessment of how the oil spill harmed larval fish and the thousands of other denizens of the gulf.

During the heat of the crisis last year, thousands of men and women flocked to the spill zone to stop the oil leak and contain its damage. A year later, the Deepwater Horizon site is strangely quiet. The last drillship pulled out in February and vanished over the horizon. On the water's surface, there are no lasting impressions of the crisis, but not so below. The wreckage of one of the world's most advanced drilling rigs lies hidden on the sea floor, as do the ecological damages that are proving so challenging to assess. ■

Mark Schrope *is as a freelancer writer in Melbourne, Florida.*

1. The Federal Interagency Solutions Group, Oil Budget Calculator Science and Engineering Team *Oil Budget Calculator: Deepwater Horizon* available at go.nature.com/tfnxjf.
2. Schrope, M. *Nature* **466,** 680–684 (2010).
3. Camilli, R. *et al. Science* **330,** 201–204 (2010).
4. Schrope, M. *Nature* **470,** 317–318 (2011).
5. Williams, R. *et al. Conserv. Lett.* doi:10.1111/j.1755-263X.2011.00168.x (2011).
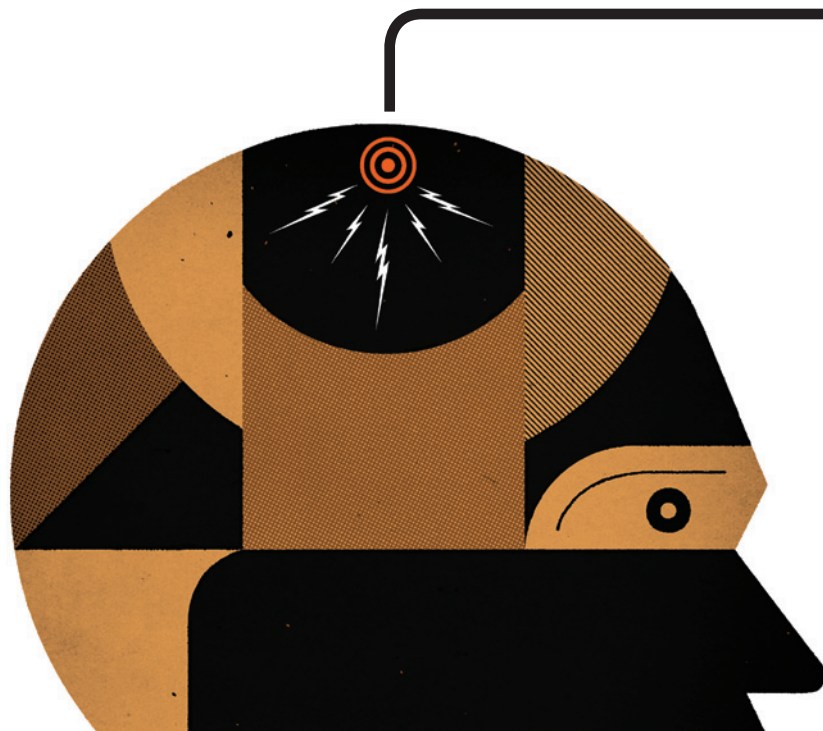
BRAIN BUZZ

BY DOUGLAS FOX

Scientists reviving a decades-old technique for brain stimulation have found that it can boost learning. So what else can be done with some wires and a nine-volt battery?

Last year a succession of volunteers sat down in a research lab in Albuquerque, New Mexico to play *DARWARS Ambush!*, a video game designed to train US soldiers bound for Iraq. Each person surveyed virtual landscapes strewn with dilapidated buildings and abandoned cars for signs of trouble — a shadow cast by a rooftop sniper, or an improvised explosive device behind a rubbish bin. With just seconds to react before a blast or shots rang out, most forgot about the wet sponge affixed to their right temple that was delivering a faint electric tickle. The volunteers received a few milliamps of current at most, and the simple gadget used to deliver it was powered by a 9-volt battery.

It might sound like some wacky garage experiment, but Vincent Clark, a neuroscientist at the University of New Mexico, says that the technique, called transcranial direct-current stimulation (tDCS), could improve learning. The US Defense Advanced Research

Projects Agency funded the research in the hope that it could be used to sharpen soldiers' minds on the battlefield. Yet for all its simplicity, it seems to work.

Volunteers receiving 2 milliamps to the scalp (about one-five-hundredth the amount drawn by a 100-watt light bulb) showed twice as much improvement in the game after a short amount of training as those receiving one-twentieth the amount of current[1]. "They learn more quickly but they don't have a good intuitive or introspective sense about why," says Clark.

The technique, which has roots in research done more than two centuries ago, is experiencing something of a revival. Clark and others see tDCS as a way to tease apart the mechanisms of learning and cognition. As the technique is refined, researchers could, with the flick of a switch, amplify or mute activity in many areas of the brain and watch what happens behaviourally. The field is "going to

explode very soon and give us all sorts of new information and new questions", says Clark. And as with some other interventions for stimulating brain activity, such as high-powered magnets or surgically implanted electrodes, researchers are attempting to use tDCS to treat neurological conditions, including depression and stroke. But given the simplicity of building tDCS devices, one of the most important questions will be whether it is ethical to tinker with healthy minds — to improve learning and cognition, for example. The effects seen in experimental settings "are big enough that they would definitely have real-world consequences", says Martha Farah, a neuroethicist at the University of Pennsylvania in Philadelphia.

Getting to this point, however, was hardly straightforward. Direct-current brain stimulation has emerged from a

long, touch-and-go history that ranges from the simply bizarre to the simply irreproducible. And for some, it still has much to prove.

The Italian scientist Jean Aldini first tried direct-current stimulation around 1800 — initially to induce movement in the corpses of recently executed felons. Later, he claimed in a paper to cure two acquaintances of the mood disorder then known as 'melancholy'. By the 1940s, many patients with depression were being given electric shocks to the temples that were strong enough to induce seizures — so-called electroconvulsive therapy. But for decades people toyed with the idea of treating mental illness with electric shocks that were much milder — 1,000 times less intense than electroconvulsive therapy.

WEAK ORIGINS

In 1964, Joe Redfearn, a psychiatrist at Graylingwell Hospital in Chichester, UK, applied some promising results in rats directly to humans, delivering weak currents — of 50–250 microamps — to the scalps of volunteers. He reports that the volunteers became talkative, even giggly, when current was run in one direction, but withdrawn when it ran the other way[2]. He gave the 'giggly' treatment to 29 patients with depression and claims that half of them improved[3]. But no one could replicate his results, and the technique was abandoned.

In retrospect, several factors seem to have undermined his work. Among them, Redfearn used currents ten times lower than in modern tDCS — perhaps because he had no way to measure how much electricity was actually reaching his patients' brains. Within a few decades, however, the necessary methods would become available, notably as researchers began to study brain activity induced by transcranial magnetic stimulation (TMS). In TMS, a magnetic coil running at thousands of volts is positioned just outside the head, leading to electrical surges inside the brain that can be precisely measured with external electrodes. Alberto Priori, a neuroscientist now at the University of Milan in Italy, showed in the 1990s that tDCS increased the effectiveness of TMS. He stimulated the motor cortex of volunteers for seven seconds with a direct current of 0.5 milliamps, then started hitting the area with short bursts of TMS.

The assumption was that if tDCS made neurons more responsive, then more of those neurons would respond when TMS was subsequently applied. It turned out to be true — volunteers who got a short pulse of direct current had a larger response to TMS. But when Priori presented his results in 1993, colleagues doubted that the electricity was penetrating the skull. It took him until 1998 to convince reviewers that his results were bona fide[4].

Michael Nitsche, a clinical neurologist at the University of Göttingen in Germany, was intrigued by the published findings. He had been experimenting with TMS to treat epilepsy at the time — but the equipment is unwieldy and expensive, and its effects on brain activity were too brief to help patients. Nitsche, a recent graduate at the time, and his supervisor, Walter Paulus, spent a year fiddling with tDCS. Their interest alarmed their colleagues. "It's fucking dangerous," Nitsche recalls being told. "You should stop this immediately." Nitsche managed to get his studies approved by university ethics boards, but a shortage of volunteers willing to have their brains zapped often forced him to experiment on his father, his sister and himself.

In 2000, Nitsche and Paulus published a paper[5] showing that up to five minutes of weak current — around 1 milliamp — on the human scalp renders the motor cortex more responsive to signals for several minutes after the electricity is shut off. Like Priori, he used TMS to measure the effects.

Nitsche and others have begun to clarify how tDCS works. Physiological studies indicate that direct current creates an electric field in brain tissue that changes the voltage across the neuronal membranes. 'Anodal' stimulation, in which electrons flow into the electrode on the head, pulls neurons a few millivolts towards 'depolarization', making them more likely to fire when signals arrive from other cells. 'Cathodal' stimulation, in which electrons flow out of the electrode on the head, has the opposite effect, 'hyperpolarizing' neurons and making them less responsive to signals from other cells.

Effects seen after the electricity is shut off can last for an hour or so and seem to arise from a second mechanism. Pharmacological evidence suggests that the current increases the expression of proteins called NMDA receptors at the

synapses, the connections between neurons. This heightens the plasticity of brain tissue — leaving it in a temporary state somewhat like wet clay, in which it is more apt to reshape its synaptic connections in response to stimuli, such as when learning a video game.

Researchers are exploring the ways in which this wet-clay state can be exploited. In a 2009 study[6], Leonardo Cohen at the National Institute of Neurological Disorders and Stroke in Bethesda, Maryland, showed that tDCS improved people's ability to learn a simple coordination exercise — and that the improvement was still apparent three months after the experiment ended. Such results have led to an interest in stroke rehabilitation strategies. Small trials by Cohen, Nitsche, and others have shown improved recovery of hand function when tDCS is used this way (see 'Wired up').

Another group of researchers, led by Felipe Fregni of the Berenson-Allen Center for Noninvasive Brain Stimulation in Boston, Massachusetts, and Paulo Boggio of Mackenzie Presbyterian University in São Paulo, Brazil, is experimenting with tDCS as a way to treat depression. Several small trials done by this group and others suggest that a few sessions of tDCS to a part of the brain called the dorsolateral prefrontal cortex can improve mood for several weeks.

RISKY APPROACH

In 2007, Boggio and Fregni reported that applying tDCS to the same region can make people less likely to take risks[7]. They asked healthy university students to play a game in which they press a computer key to pump air into a cartoon balloon. The more they pump, the more virtual money they earn — but if the balloon bursts, they lose all their winnings. People treated with tDCS were less willing to push their luck. The results may be generalizable to addictions, in which people lack "inhibitory control", says Boggio. In 2008, he and Fregni published three studies[8–10] showing that stimulation of the dorsolateral prefrontal cortex blunted cravings for alcohol, cigarettes and sweets when people later watched videos in which these were being consumed. They hope, eventually, to test the same technique in a clinical trial for smoking cessation.
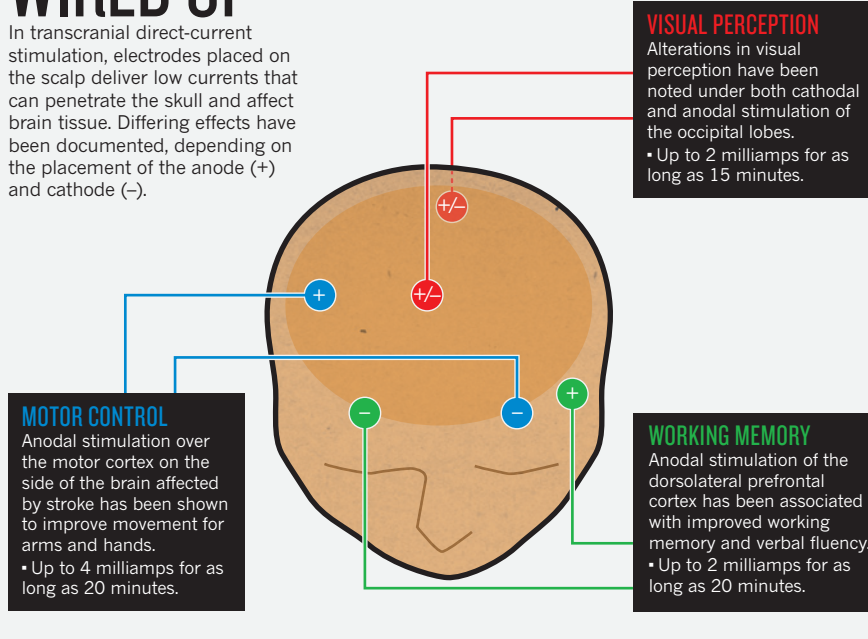
> "There has been a lot of hokey stuff, and it affects the credibility of the entire field."

For a method that has seen its ups and downs, these results are encouraging. "There has been a lot of hokey stuff, frankly, and it affects the credibility of the entire field," says Marom Bikson, a biomedical engineer at the City College of New York. But in contemporary mechanistic studies to optimize tDCS, he says, "people have been much more careful".

Not everyone is convinced that the
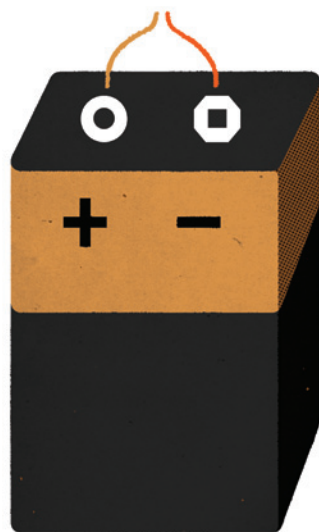
WIRED UP

In transcranial direct-current stimulation, electrodes placed on the scalp deliver low currents that can penetrate the skull and affect brain tissue. Differing effects have been documented, depending on the placement of the anode (+) and cathode (–).

VISUAL PERCEPTION
Alterations in visual perception have been noted under both cathodal and anodal stimulation of the occipital lobes.
• Up to 2 milliamps for as long as 15 minutes.

MOTOR CONTROL
Anodal stimulation over the motor cortex on the side of the brain affected by stroke has been shown to improve movement for arms and hands.
• Up to 4 milliamps for as long as 20 minutes.

WORKING MEMORY
Anodal stimulation of the dorsolateral prefrontal cortex has been associated with improved working memory and verbal fluency.
• Up to 2 milliamps for as long as 20 minutes.

disappointments are over, though. Helen Mayberg, a clinical psychiatrist at Emory University in Atlanta, Georgia, has been experimenting with the use of deep-brain stimulation (DBS), in which electrodes are placed deep in the brain, to treat depression. She is excited about the non-invasiveness of tDCS but points out that the trials conducted to date have been short-term. The real questions, she says, will be: "How do you use it chronically, and what kinds of rebounds and relapses are there?" Boggio and others have begun a clinical trial to answer this question, in which patients being treated with tDCS for depression will be observed for up to six months.

TARGET PRACTICE
Still, the stimulation from tDCS is less focused than that from TMS or DBS. Its effects on neurons also drop off rapidly a few centimetres below the scalp, putting some important medical targets out of reach.

Bikson has designed a more refined version of tDCS that he hopes will address these shortcomings. Instead of one electrode, he places five on the head in an X configuration. The one in the centre pushes current in the desired direction and the four around it siphon off excess current that would otherwise spread and activate wider brain areas. The configuration could, he says, allow for slightly higher currents that would penetrate deeper into the brain in more focused areas. Such innovations might even help to persuade companies to invest in clinical trials. Currently, Cohen says, no one stands to gain enough return from therapies that can be administered using just US$1,000 worth of off-the-shelf equipment.

Aside from treatment, tDCS is also receiving attention for its potential to enhance the minds of healthy people. In addition to Clark's work showing enhanced ability to see concealed threats, other studies with tDCS have shown improvements in working memory[11], word association[12] and complex problem-solving[13]. Most of these studies address scientific questions — but one neuroscientist unabashedly aims to boost the brains of healthy people.

Allan Snyder, director of the Centre for the Mind at the University of Sydney in Australia, hopes to develop "a thinking cap", a tDCS device that corporate executives or advertising copywriters might use to bump up their creativity before walking into a brainstorming meeting. Snyder is cagey about how far he is in product development — but his latest demonstration, published this February[14], garnered plenty of attention. Snyder claims to have boosted people's flair for sudden insight by stimulating their anterior temporal lobes. People who received tDCS were two to three times more likely than those receiving sham stimulation to solve a creativity problem in which they raced against the clock to spell out maths equations with matchsticks.

The jury is still out on whether these results will translate into real-world benefits. Nitsche says that it will be harder to improve cognition in young, healthy people — whose minds are theoretically already optimized — than in elderly people or those with addictions,

> "With wires and batteries and home hobbyists trying to run electricity through their heads, somebody could get hurt."

for instance. "I wouldn't say it wouldn't be possible," says Nitsche. "But things might be a little more complicated."

That's not stopping some people from trying it at home. Discussions are already appearing on the Internet: buy a 9-volt battery, some wire and a resistor, and you're theoretically there. One person, hoping to improve his concentration, was alarmed by the flashing lights he experienced — a commonly reported side effect, along with burning or itching at the site of the electrode. "I probably won't be doing this again," he said in a message posted online. Another wrote in an online patients' forum that the tDCS treatments he was giving to his wife were alleviating her chronic pain. Safety is an important issue. "With wires and batteries and home hobbyists trying to run electricity through their heads, somebody could get hurt," says Farah.

And wider adoption raises ethical concerns similar to those that surround mind-enhancing drugs such as Adderall and Modafinil, which some students take as study aids. Students might secretly 'electrodope' with tDCS before a university entrance exam to inflate their scores. Ethicists worry that this will give some an unfair advantage or create a culture in which people feel pressured to use such devices. None of the studies published so far have shown a type of mind-sharpening that would help in such exams, says Farah, but that might simply be a matter of targeting the right brain areas. "It would not surprise me" if such effects were possible, she says.

Overall, though, the optimism among tDCS's believers remains high. Although it has generated some disappointments, many are convinced that the present buzz is warranted. "Sometimes in the history of medicine you have to try again after one century or so," says Priori. "You use a novel technical device, and you succeed where somebody else failed." ∎

Douglas Fox is a freelance writer in Northern California.

1. Clark, V. P. *et al. NeuroImage* advance online publication doi:10.1016/j.neuroimage.2010.11.036 (2010).
2. Lippold, O. C. J. & Redfearn, J. W. T. *Br. J. Psychiatr.* **110,** 768–772 (1964).
3. Redfearn, J. W., Lippold, O. C. & Costain, R. *Br. J. Psychiatr.* **110,** 773–785 (1964).
4. Priori, A., Berardelli, A., Rona, S., Accornero, N & Manfredi, M. *Neuroreport* **9,** 2257–2260 (1998).
5. Nitsche, M. A. & Paulus, W. *J. Physiol.* **527,** 633–639 (2000).
6. Reis, J. *et al. Proc. Natl Acad. Sci. USA* **106,** 1590–1595 (2009).
7. Fecteau, S. *et al. J. Neurosci.* **27,** 6212–6218 (2007).
8. Fregni, F. *et al. Appetite* **51,** 34–41 (2008).
9. Boggio, P. S. *et al. Drug Alcohol Depend.* **92,** 55–60 (2008).
10. Fregnn, F. *et al. J. Clin. Psychiatr.* **69,** 32–40 (2008).
11. Ohn, S. H. *et al. Neuroreport* **19,** 43–47 (2008).
12. Cerruti, C & Schlaug, G. *J. Cogn. Neurosci.* **21,** 1980–1987 (2009).
13. Dockery, C. A., Hueckel-Weng, R., Birbaumer, N. & Plewnia, C. *J. Neurosci.* **29,** 7271–7277 (2009).
14. Chi, R. P. & Snyder, A. W. *PLoS One* **6,** e16655 (2011).

COMMENT

J. EVESON/FLPA/PHOTOLIBRARY

Applying liquid manure more precisely than this would be cleaner, reduce odour and emit less ammonia.

Too much of a good thing

Curbing nitrogen emissions is a central environmental challenge for the twenty–first century, argue **Mark Sutton** and his colleagues.

An analysis published this week calculates that excess nitrogen in the environment costs the European Union (EU) between €70 billion (US$100 billion) and €320 billion per year[1]. It is the first time that an economic value has been placed on the threats posed by nitrogen pollution, including contributions to climate change and biodiversity loss. To put that into perspective, this cost is more than double the value that nitrogen fertilizers are estimated to add to European farm income.

This economic valuation is just one outcome of the five-year European Nitrogen Assessment, which has drawn together 200 experts, including ourselves, to understand the sources, processes and impacts of nitrogen as a basis to inform policy.

Excess reactive nitrogen threatens the quality of air, soil and water. It affects ecosystems and biodiversity, and alters the balance of greenhouse gases. Existing climate and air-pollution policies that aim to reduce energy consumption and fossil-fuel burning, such as the Gothenburg Protocol, are helping to cap nitrogen emissions ▶

SUMMARY

● Nitrogen pollution costs the European Union between €70 billion and €320 billion per year.
● Policies should address farming, meat consumption, use of human sewage and fossil-fuel burning.
● The Gothenburg Protocol is an opportunity to further reduce emissions.
● A global inter-convention nitrogen protocol is needed.

DAMAGE COSTS OF NITROGEN POLLUTION

Agriculture and fossil-fuel burning load the environment with reactive nitrogen, affecting water, soils and air.

MAIN NITROGEN SOURCES

Legend:
- Reactive nitrogen runoff in water
- NH_3 released to air
- NO_x released to air
- N_2O released to air

Axis: Cost to European Union (billion €) — Human health, Ecosystems, Climate

Sources: Agriculture, Fossil-fuel burning, Industry, Sewage, Other
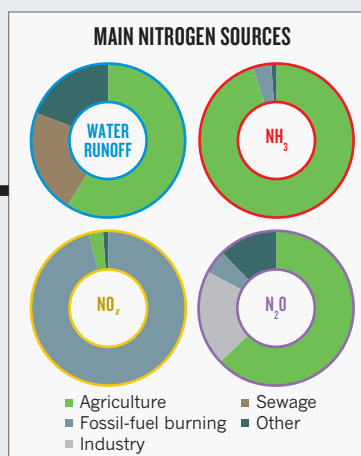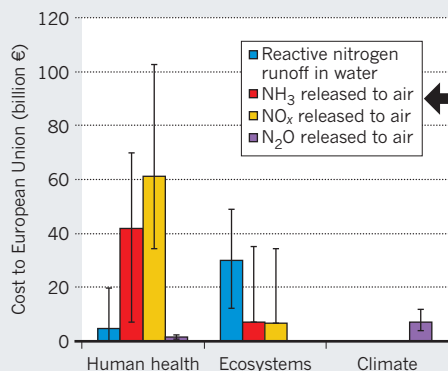
▶ from transport and industry. Together with policies on nitrates in drinking water and reductions in cattle numbers, these have led to a modest drop in European nitrogen pollution since the 1980s.

But existing policies are piecemeal, and there has been little attempt to integrate the effects of the different nitrogen threats. Nitrogen pollution poses an even greater challenge than carbon, because the element has many complex effects as it cascades through many chemical forms. There is a need for new policies that cover these concerns and encourage changes in farming, diet and even what we do with human waste.

THE DAMAGE DONE

Nitrogen can be divided into two classes. Unreactive nitrogen (N_2) makes up 78% of Earth's atmosphere. Reactive nitrogen includes every other form of the element, including nitrogen oxides (NO_x), nitrous oxide (N_2O), ammonia (NH_3) and nitrate (NO_3). All biological systems need reactive nitrogen, but historically it has been in short supply. Until the end of the nineteenth century, the main agricultural source was fixation of N_2 by symbiotic bacteria in legumes planted for that purpose, combined with careful recycling of the limited amount of nitrogen in manure.

By 1900, there was a worrying shortage of reactive nitrogen, as fertilizers were needed to feed a growing population and nitrogen-based explosives were wanted for weaponry. The solution came in 1908 with the Haber–Bosch process, allowing cheap ammonia to be made from N_2 and energy on an industrial scale. This process was so successful that, within a century, human production of reactive nitrogen had more than doubled global rates of nitrogen fixation. Without it, around half of humanity would not be alive[2]. As an intentional strategy to address the world's shortage of reactive nitrogen, the Haber–Bosch process is arguably the greatest single experiment ever made in geoengineering.

But the mass production of reactive nitrogen has come with costs. Nitrogen-based explosives killed 100 million people in armed conflicts during the twentieth century[2]. Their use in mining has opened up fossil-fuel reserves, the burning of which has released carbon dioxide and more reactive nitrogen into the atmosphere. About half of the nitrogen added to farm fields in Europe ends up as pollution, or is wasted by denitrification back to N_2 (ref. 1).

In water, excess reactive nitrogen causes algal blooms that can kill fish, and nitrate in drinking water can harm human health, increasing the risk of bowel cancer. About 80% of European fresh waters exceed a threshold for high risk to biodiversity of 1.5 milligrams of nitrogen per litre. Combined with the effects of too much phosphorus, the resulting proliferation of algae in coastal areas gives new meaning to the idea of 'green' bathing.

In the air, reactive nitrogen adds to particulate matter and ground-level ozone, created when nitrogen oxides react with organic compounds, causing respiratory and cardiovascular disease. The effects of airborne particles take six months off the life expectancy of at least half of all Europeans. In forests, ammonia pollution encourages the growth of algal slime that can suffocate tree-living plants such as mosses and lichens (see picture) to such an extent that it would be easy to think this were natural in agricultural areas of northwest Europe. The assessment estimates that NH_3 and NO_x emissions have reduced forest biodiversity by more than 10% over two-thirds of Europe. Nitrogen deposition also threatens the ability of peatlands to store carbon by killing the bog-building moss *Sphagnum*.

Importantly, we now have a first estimate of how the different climate warming and cooling effects of nitrogen add up across Europe. Nitrogen tends to warm the planet by forming N_2O and ground-level ozone, both of which are powerful greenhouse gases.

At the same time, nitrogen emissions tend to cool the planet: by reducing the atmospheric life-time of methane; by forming particulate matter that reflects light back into space; and by acting as a fertilizer, increasing the growth of forests. Overall, the assessment finds that these effects tend to balance out. Efforts need to focus on reducing the warming effects, while recognizing that the adverse effects of particulate matter and nitrogen deposition on human health and biodiversity may more than outweigh their climate benefits.

COSTS AND BENEFITS

It is hard to tally up all the costs and benefits of reactive-nitrogen production. The direct and indirect effects influence everything from population growth and energy production to the manufacture of nitrogen-based products such as nylon, polyurethane and hydrazine rocket fuel.

The benefits of nitrogen can be seen from its largest use: agricultural fertilizers. Manufactured fertilizer adds 11 million tonnes of reactive nitrogen annually to fields in the EU. This produces a direct benefit to European farmers, in terms of crops grown, of €20 billion to €80 billion per year, when the long-term benefits are included[1]. On top of that, biological fixation and recycled sources of nitrogen — including plant residues, animal manures and atmospheric deposition — add 17 million tonnes, giving a total direct benefit of €25 billion to €130 billion, even before value is added during the food-supply chain.

But half of the nitrogen in fertilizers and manures is lost to the surrounding environment. In economic terms, this amounts to a loss of potential benefits to farmers of €13 billion to €65 billion per year. On these grounds alone, there is a strong case for using nitrogen more efficiently.

> **"The effects of airborne particles take six months off the life expectancy of at least half of all Europeans."**

These numbers, however, are much lower than the estimated cost of the damage from reactive-nitrogen emissions. The assessment gauges these costs by looking at society's readiness to pay for longer and healthier life, for nature restoration, or for climate stability. The grand total of €70 billion to €320 billion per year is equivalent to 1–4% of the average disposable per capita income of European citizens.

Of the total cost of damage from reactive-nitrogen emissions, 75% comes from the effects of NO_x and NH_3 on human health and ecosystems. The effect of N_2O emissions on climate warming accounts for only around 5%. Although N_2O has recently been heralded as the main cause of stratospheric ozone depletion[3], this represents only 1% of the damage costs. Climate change and ozone

thinning are important, but the threats to health and ecosystems are an even stronger argument for taking action on nitrogen.

Clearly nitrogen is one of the major environmental challenges of the twenty-first century.

Although better combustion and energy technologies can reduce nitrogen emissions from fossil fuels, further efforts must curb the increasing use of energy and transport. We also need a radical rethink of sewage systems, which waste reactive nitrogen, using yet more energy to turn it back to N_2. The technology is available to recycle sewage nitrogen, but it needs development[4]. Such a major change is certainly possible, as shown by nineteenth-century Paris, where more than half of the nitrogen in sewage was recycled, including in the manufacture of ammonium sulphate fertilizer on an industrial scale[1]. One approach would be to use sewage to make biogas, leaving a useful residue rich in nitrogen and phosphorus.

But the largest challenges are to manage nitrogen better in agriculture and to moderate Europeans' consumption of animal protein. Amazingly, livestock consume around 85% of the 14 million tonnes of nitrogen in crops harvested or imported into the EU; only 15% is used to feed humans directly[1]. European nitrogen use is therefore not primarily an issue of food security, but one of luxury consumption. If Europeans obtained all their protein from plants, only 30% of the crops grown currently would be needed, reducing nitrogen fertilizer inputs and the associated pollution by 70%.

DRIVING CHANGE

The average EU citizen now eats much more meat and milk than is needed for a healthy diet[1]. Reducing meat and dairy consumption could therefore benefit the health of many people and help protect the environment at the same time. Recognizing this, in 2009 some of the authors of this piece developed the 'Barsac Declaration', through which many of the assessment authors have committed to fostering the 'demitarian diet'[5], making it easier to choose half-portions of meat. Such initiatives have a huge potential to change dietary aspirations in a world where per-capita meat consumption is rapidly increasing.

There are many ways to reduce nitrogen losses from agriculture. Improving the production potential of crops and livestock can help, and covers on manure storage tanks and low-emission techniques for spreading fertilizers and manure are essential. At present, most liquid manure is spread with a crude spraying method that dirties the

crop and maximizes NH_3 emissions. Applying manure in bands cuts emissions by 30–60%, decreases odours, leaves the crop clean and reduces the farmer's fertilizer bill. By improving nitrogen-use efficiency in this way, measures to decrease NH_3 emissions turn out to be central to reducing overall N_2O emissions.

Experience shows that regulatory or financial drivers are needed to ensure such technologies are used. Changes need not be uneconomic. The Netherlands and Den-

Ammonia can kill tree-living plants (left), replacing them with algal slime.

mark have required low-emission techniques for more than a decade, yet livestock farming in these countries remains among the most competitive in Europe.

At present, no single United Nations (UN) convention can handle all the threats posed by reactive nitrogen. An inter-convention protocol is needed. This could, for example, link the UN Framework Convention on Climate Change and the Convention on Biological Diversity with the Air and Water Conventions of the United Nations Economic Commission for Europe (UNECE).

In the short term, there are already opportunities on the table for Europe. The UNECE Air Convention — also known as the Convention on Long-range Transboundary Air Pollution — is currently renegotiating its Gothenburg Protocol, which caps national emissions of air pollutants. Like the update of the Kyoto Protocol, this is happening slowly: talks began in 2007. Further reductions in NO_x emissions will surely be agreed. The big question is how far the parties will commit to reducing NH_3 emissions.

The air convention's Task Force on Reactive Nitrogen has recently shown that preventing NH_3 release is much cheaper than previously estimated, especially once the fertilizer savings are considered. Scaling up these numbers shows that cutting NH_3 emissions across the EU by 20% would cost less than €500 million per year[6]. On the basis of total NH_3 damage costs of €15

billion to €105 billion per year (see graphic), the environmental benefits of mitigation are therefore around 20 (range of 6 to 42) times the costs, even without counting the other benefits this gives in reducing N_2O emissions and nitrogen runoff.

Given the political sensitivity of the agricultural sector, however, the case for reducing NH_3 emissions is likely to depend on demonstrating the broader impact of better nitrogen management. This challenges the scientific community to continue quantifying the benefits of nitrogen mitigation. Now that a first effort has been made for Europe, a global assessment of nitrogen is urgently needed, as called for by the International Nitrogen Initiative in their 2010 Delhi Declaration[7].

At the same time, the policy community must develop more joined-up approaches. If parties to the Gothenburg Protocol, for example, are to agree serious commitments to reducing NH_3 and NO_x emissions, they will need a motivation that can only come from seeing the bigger picture. They should be jumping at the chance that reducing NH_3 and NO_x emissions could help meet their existing N_2O and NO_3 commitments. Such ideas point to a vision for better nitrogen management where air, soil, water, climate and biodiversity can all benefit. ∎

Mark A. Sutton *of the Centre for Ecology and Hydrology, Edinburgh EH26 0QB, UK,* and **Oene Oenema** *of Alterra, Wageningen 6700 AA, the Netherlands, are the co-chairs of the UNECE Task Force on Reactive Nitrogen. They are contributors to the European Nitrogen Assessment with* **Jan Willem Erisman** *at the Vrije Universiteit Amsterdam, 1081 HV Amsterdam, the Netherlands;* **Adrian Leip** *at the European Commission-Joint Research Centre, Institute for Environment and Sustainability, I-21027 Ispra, Italy;* **Hans van Grinsven** *at the PBL Netherlands Environmental Assessment Agency, 3720 AH Bilthoven; and* **Wilfried Winiwarter** *at the International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria.*
e-mail: ms@ceh.ac.uk

1. Sutton, M. A. *et al.* (eds) *The European Nitrogen Assessment* (Cambridge Univ. Press, 2011); available at http://go.nature.com/5n9lsq
2. Erisman, J. W. *et al. Nature Geosci.* **1,** 636–639 (2008).
3. Ravishankara, A. R. *et al. Science* **326,** 123–125 (2009).
4. Magid, J. *et al. Ecol. Eng.* **28,** 44–54 (2006).
5. *The Barsac Declaration: Environmental Sustainability and the Demitarian Diet* (2009); available at http://go.nature.com/f5e1lh
6. Amann, M. *et al. CIAM report 1/2011* (IIASA, Austria, 2011).
7. *Delhi Declaration* (2010); available at http://go.nature.com/mauqpr

Existing plans to contain and clean up any oil spill that might occur in the Arctic are woefully inadequate.

A frozen hell

A year after the oil blowout in the Gulf of Mexico, **Jeffrey Short** and **Susan Murray** call for action to prevent an even more nightmarish scenario: a spill in the Arctic.

In a world of rising oil prices, there is pressure to drill for oil wherever it may be found — including 'frontier' areas such as the Arctic Ocean. With an estimated 80 billion barrels, or 4% of all the oil that could be recovered by conventional means, Arctic waters loom large in industry thinking. But extracting oil there comes with immense risks, thanks to the region's remoteness and harsh conditions, a dearth of experience in offshore drilling there and a rudimentary understanding of the marine ecosystem. On the anniversary of the Deepwater Horizon oil spill, itself the result of oil and gas exploration under extreme conditions, we outline the disastrous consequences of an Arctic spill, and consider how to prevent it.

Much is at stake. One area marked for development, the Chukchi Sea (see map), is among the world's most productive ocean areas. Throughout the Arctic, algae support a food web that includes clams, crabs, fish and marine mammals; millions of seabirds migrate there and indigenous communities rely on a healthy marine ecosystem for their subsistence way of life.

Warming conditions have expanded ice-free areas, enabling more fishing, shipping and oil and gas extraction. Development is most advanced in Alaska, which is thought to hold by far the largest portion of Arctic offshore oil. Onshore exploration there during the 1960s led to discovery of the 25-billion-barrel Prudhoe Bay oil field in 1968. The first production in the Beaufort Sea began from a gravel island at the Endicott field in 1987. In the US Arctic, ocean leases worth more than US$7 billion have been sold since 1980, leading to approximately 35 exploration wells, the discovery of a 206-million-barrel field (Northstar) and continued industry interest.

AT THE MERCY OF THE ELEMENTS

The Norwegian Sea's Snøhvit gas and condensate field began production in 2006, and plans are in place to develop fields off the Lofoten Islands. Exploration is also planned this year for the east and west coasts of Greenland, and the Arctic continental shelves of Russia and Canada.

Most of the Arctic offshore oil lies under less than 500 metres of water; the Deepwater Horizon rig was in 1,500 metres of water. But drilling is still hard. Along the Alaskan coast, nearly constant winds average about 20 kilometres per hour and can exceed 100 kilometres per hour, causing high seas. Frequent fogs and storms reduce visibility. Drilling rigs can be engulfed in ice floes up to a metre thick. All this, along with the distances to markets and supporting infrastructure, makes production extremely costly, requiring giant reservoirs of oil to justify initial investments. Giant reservoirs take a long time to drain, increasing the chance of a spill over a site's lifetime.

On the basis of extrapolations from more temperate climes, the US Department of the Interior puts the risk of a 5,000-barrel marine spill (one-fifth the size of the *Exxon Valdez*) at 40% for the Chukchi field over its operational life. The agency dismissed the risk of more catastrophic spills, but the Deepwater Horizon blowout has led some to re-appraise that view. Although the risks of low-probability, high-impact events are difficult to estimate, it is sobering that each major marine oil production area in the United States has seen at least one catastrophic spill: the 1969 blowout of a drilling rig off the coast of Santa Barbara, California; the 1989 *Exxon Valdez* tanker spill in Alaska; and the 2010 Deepwater Horizon blowout in the Gulf of Mexico.

OCEAN/CORBIS

Stopping, mitigating and cleaning up an Arctic spill would all be seriously impeded. In the Alaskan Arctic, the nearest US Coast Guard base is more than 1,500 kilometres away, and airstrips are small, few and scattered. Fog and snowstorms could ground workers for weeks at a time. Skimming or burning rarely removes more than a small fraction of the oil released by a major spill (8% for the 1989 *Exxon Valdez* spill; 10% for Deepwater Horizon), even when skimmers, booms and dispersants are abundant and nearby. When the Icelandic cargo ship *Godafoss* ran aground on the southern Norwegian coast in February, ice and weather prevented response teams from keeping a few hundred tonnes of heavy fuel oil from killing hundreds of birds and contaminating the shorelines of a nearby marine park. Shell has plans for "unprecedented oil spill response capabilities, including dedicated at-site vessels, a containment system and relief well rigs", according to the company. These are no doubt state-of-the-art and welcome commitments, but their effectiveness in Arctic conditions has yet to be tested. The record of oil-spill response elsewhere, under less challenging conditions, fails to inspire confidence.

BLACK DEATH

If oil does spill, a small fraction will dissolve into sea water, a larger fraction will evaporate and most will be slowly oxidized by microbes. Studies from an experimental oil spill in Baffin Island in 1981 and the 1989 *Exxon Valdez* spill show that microbial oxidation can remove much of the oil within a year. But buried masses of oil can persist for decades, sea ice can envelop oil and transport it considerable distances, and some oil might sink, contaminating seafloor communities. A blowout during autumn would spill among growing ice floes, spreading contamination further than it could be tracked and concentrating oil in the ice holes through which marine mammals breathe.

Evaluating the effects of oil discharged into these ecosystems is especially challenging because we know so little about them. In the 1970s and 80s, the Outer Continental Shelf Environmental Assessment Program provided information for the seas around southern Alaska, but less effort was devoted to the Arctic. Our knowledge of the Arctic's species is still patchy, and there is sparse monitoring for basic physical factors such as the speed and direction of winds and ocean currents.

Perhaps the greatest damage from a spill would be to the region's indigenous peoples. Fears about the safety of subsistence foods may erode hunting skills, and cause the younger generation to question the knowledge and wisdom of its elders. This happened to the Alutiiq peoples following the *Exxon Valdez* oil spill; oily clams and fish deterred many from their traditional hunting and gathering for years.

Environmental concerns have motivated conservation groups to insist on regulatory oversight, with some success. Court rulings and administrative actions have undermined leasing decisions and postponed drilling in the Chukchi and Beaufort seas. Citing the environmental and economic consequences of a spill, the Norwegian government recently decided to delay

RUSSIA

Chukchi Sea leases sold, exploration contested

ALASKA

Endicott - since 1986
Northstar - since 2001
Beaufort Sea leases sold, exploration contested
Mackenzie Delta discontinued

North Pole ●

Sverdrup basin discontinued

CANADA

Barents Sea

Prirazlomnoye platform being built

Goliat production planned for 2013

Snøhvit gas since 2006

Lofoten Islands development suspended

GREENLAND

Western Greenland exploration planned for summer 2011

NORWAY

Eastern Greenland exploration plans in early stages

● Oil exploration
● Oil production
▮ Oil and gas reserves

development of the Lofoten field.

A recent report by the US president's oil-spill commission, *Deep Water: The Gulf Oil Disaster and the Future of Offshore Drilling*, states that for activities to move forward in the Arctic there should be containment and response plans at every stage, the US Coast Guard and oil companies should be able to deal with an accident, and Congress should provide the resources to ensure that the Coast Guard has sufficient presence. Although laudable, literal adherence to these principles would halt offshore oil development immediately and indefinitely. We know that containment and response plans are woefully inadequate, and the cost of increasing the currently negligible US Coast Guard presence in the Arctic is immense and unlikely to be funded in the current budgetary climate.

PREPARING FOR THE WORST

Recognizing that the development of Arctic oil fields is likely, we recommend three guiding principles. First, greater investment in a science programme to provide an understanding of the environment in these new petroleum-producing provinces. Studies in the US Arctic have been useful, but they have been narrow and disjointed. Information gaps remain, including the locations of biologically important 'hot spots' that deserve priority protection.

Second, industry should be held to strict standards. A large spill in the Arctic could not be contained or mitigated, and we should stop pretending otherwise. Regulations and their enforcement should be strong enough to deter companies from skirting them. And oversight of industry should be shared with third-party citizen groups such as the Regional Citizen's Advisory Councils in the United States.

Finally, governments should consider how to manage exploration in ever-more challenging environments. The benefits of oil go mainly to shareholders, whereas the public bears the risks. Possible solutions include raising or removing liability limitations for spills, and requiring oil companies to issue performance bonds — advanced security deposits that cover the cost of a catastrophe.

The public is becoming increasingly aware of the need for caution in developing Arctic resources. The recently adopted US Arctic fishery management plan, for example, calls for a better understanding of the Arctic marine ecosystem before commercial fisheries are authorized. We can still minimize the impacts of oil development, but only if we avoid placing economic benefit above all else. The precedent set by the United States will strongly influence standards adopted elsewhere. ∎

Jeffrey Short *runs JWS Consulting out of Juneau, Alaska, USA, and previously worked for the US National Oceanic and Atmospheric Administration.* **Susan Murray** *is senior Pacific director of the advocacy group Oceana, Juneau, Alaska 99801, USA.*
e-mails: jwsosc@gmail.com;
smurray@oceana.org

An etching from 1705 showing a lamb-to-human blood transfusion.

MEDICINE

Blood feud

A history of early transfusions mixes experiment and ethics with Anglo–French rivalry, finds **W. F. Bynum**.

Modern blood transfusion is a wondrous mix of science and technology. It hinges on Karl Landsteiner's discovery in 1901 of the major human blood groups, which earned him a Nobel prize. He was also involved in the elucidation of the rhesus factor, important for understanding blood incompatibilities. Before Landsteiner, transfusion was a lottery at best, and fatal at worst.

In the decades before Landsteiner, a few intrepid physicians occasionally resorted to transfusion, especially in obstetric cases in which the woman's loss of blood could be treated with a donation from her husband. It was always touch and go. But at least these physicians were performing human-to-human transfusion. Two centuries earlier,

animals, usually lambs, were used as donors.

In *Blood Work*, medical historian Holly Tucker looks at the beginnings of transfusion in the seventeenth century. Adding material from her own archival research to the standard historical account, she fleshes out the start of physiological experimentation and examines historical attitudes to blood. The result is a page-turning insight into early scientific attitudes and disputes over priority.

In the 1660s, she explains, the fellows of the newly established Royal Society of London began to transfer blood from one animal to another. This was part of investigations into the heart, blood, circulation and respiration, following William Harvey's seminal description of the circulatory system in 1628. The British group included Christopher Wren, Robert Hooke, Robert Boyle and Richard Lower, and their experiments were communicated throughout Europe by the secretary of the Royal Society, Henry Oldenburg, and published in the society's *Philosophical Transactions*.

Meanwhile, the French Academy of Sciences opposed transfusion, so the main innovator in France was a marginal but ambitious physician, Jean-Baptiste Denis. He, too, began with animal-to-animal work, but quickly moved on to transferring blood between animals and humans. At stake was more than whether foreign blood was curative — people believed that the characteristics of the chosen animal might alter human personality.

The docile lamb was the donor species most often used, with its religious overtones (Agnus Dei, or Lamb of God). "The blood is the life," the Bible tells us, and seventeenth-century natural philosophers endowed that phrase with its full theological significance. Consequently, these were not simply experiments of curiosity; they were aimed at uncovering deeper meanings. That said, one of Denis's early subjects, a butcher, apparently took away his exsanguinated donor to roast.

The early human transfusion patients in both countries were generally treated for what was perceived to be lunacy or other psychiatric disabilities. The first English subject was an eccentric clergyman who liked to converse in Latin. He survived an infusion of about a third of a litre of lamb's blood.
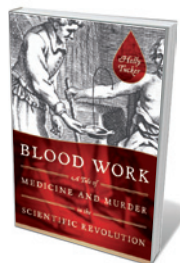
Blood Work: A Tale of Medicine and Murder in the Scientific Revolution
HOLLY TUCKER
W. W. Norton: 2011.
304 pp. $25.95

⟳ **NATURE.COM**
For a retrospective review of Fernel's *Physiologia*:
go.nature.com/83sfod

Although it did not cure his language preference, he was less agitated afterwards. So the procedure was thought to hold enough promise to be repeated a couple of weeks later.

Hopes were dashed when Denis transfused an agitated servant, Antoine Mauroy, with some calf's blood. The first two transfusions seemed to calm him. A third, insisted on by Mauroy's wife, was abandoned when the patient had a series of seizures. The next morning, Mauroy was dead, and was taken away for burial before Denis could perform an autopsy. The case came to trial, and although Denis was exonerated, Mauroy's wife was implicated in poisoning her husband. The results of her trial are lost, but a year later the French parliament prohibited transfusion of blood into humans. The English also lost heart until human-to-human transfusions were gingerly begun in the 1820s.

Tucker makes a reasonable (if circumstantial) case that one of Denis's medical opponents, Henri-Martin de la Martinière, supplied Madame Mauroy with poison to dispose of her violent husband, also putting paid to Denis's ambitions. De la Martinière was a colourful surgeon with a dark past involving pirates; he wrote vitriolic pamphlets against Denis and transfusion.

Tucker uses the competition between the French and English scientific societies as a window onto international rivalries during the scientific revolution. Claims over priority were at stake — being the first to successfully carry out a procedure mattered a great deal. In Paris, there was the additional tension between the Academy of Sciences, sanctioned by King Louis XIV, and the private groups that it effectively replaced. The fact that the French and English were at war during the 1660s adds spice to the story, as do London's 1665 plague outbreak and the Great Fire in 1666, which disrupted the early meetings of the Royal Society.

> *"Competition between the French and English scientific societies gives a window onto international rivalries during the scientific revolution."*

Tucker seeks to expose the passions and the pain behind physiological experimentation. Her narrative reads like a novel, marred only by occasional errors of fact and lapses into sensationalism. *Blood Work* is a powerful reconstruction of what has often been relegated to a minor episode within early modern science and medicine. ∎

W. F. Bynum *is emeritus professor of medical history at University College London, UK.*
e-mail: w.bynum@ucl.ac.uk

Books in brief
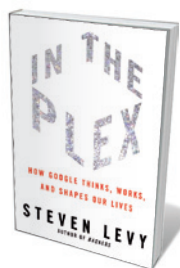
In The Plex: How Google Thinks, Works, and Shapes Our Lives
Steven Levy SIMON & SCHUSTER *432 pp. $26 (2011)*
Expanding beyond its early dominance of the search-engine market, Google has shifted gears many times since it began as a small start-up firm in Silicon Valley, California. Yet the company clings to its image as a creative hub. Technology writer Steven Levy gives an upbeat account of life inside the Googleplex campus, where hand-picked workers are able to devote up to 20% of their time to self-generated projects. He explores the next frontiers for the company, such as cloud computing and social networks, and examines its controversial decision to enter China.
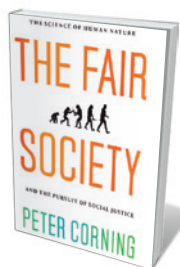
The Fair Society: The Science of Human Nature and the Pursuit of Social Justice
Peter Corning UNIVERSITY OF CHICAGO PRESS *256 pp. $27.50 (2011)*
Evolution is often reduced to the survival of the fittest, yet that does not mean we should trample on others to get ahead. We have an innate sense of fairness, argues complex-systems biologist Peter Corning in his book, which draws on our evolutionary history and the science of human nature. For the benefit and well-being of all, he proposes that society should adjust its political and economic priorities toward fairness. We should adopt a new 'biosocial' contract, which promotes the principles of equality, equity and reciprocity.

The Great Sperm Whale: A Natural History of the Ocean's Most Magnificent and Mysterious Creature
Richard Ellis UNIVERSITY PRESS OF KANSAS *432 pp. $34.95 (2011)*
Just as the sperm whale inspired Herman Melville's 1851 book *Moby Dick*, it has long fascinated writer Richard Ellis. Here he devotes a whole volume to the giant creature, which he notes has stalked the chapters of his other books on marine life, from giant squid to tuna. Using his own elegant illustrations, Ellis discusses the sperm whale's evolution and biology, its migrations, diet and breeding. He also considers its impact on myths about sea monsters, and the whaling that has decimated its numbers over two centuries.
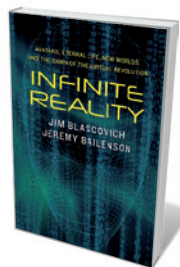
Infinite Reality: Avatars, Eternal Life, New Worlds, and the Dawn of the Virtual Revolution
Jim Blascovich and Jeremy Bailenson WILLIAM MORROW *304 pp. $27.99 (2011)*
Online environments are becoming ever more convincing and pervasive. Psychologists Jim Blascovich and Jeremy Bailenson ask how our brains cope with virtual reality. Describing the emerging technologies and what they say about us, the authors point out how our interactions are mostly driven by age-old impulses to search for new experiences and deeper perspectives on ourselves. They argue that in that sense, virtual reality is just an extension of humanity.
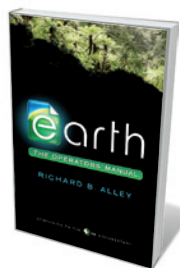
Earth: The Operators' Manual
Richard B. Alley W. W. NORTON *479 pp. $27.95 (2011)*
In a book to accompany a PBS documentary, climate scientist Richard Alley charts our expansive need for energy sources and the damage that fossil fuels are wreaking on our planet. Setting out the scientific facts clearly for the layperson rather than pushing particular solutions, he explains how we have come to be dependent on coal, oil and gas; points out the impact of greenhouse-gas emissions; and details the choices that will eventually have to be made over alternative forms of energy production.

NEUROETHICS

The origins of morality

Our values may have biological roots, finds **Adina Roskies**, but we still need philosophy.

Morality isn't just what your father, priest or favourite philosopher told you. It is a natural phenomenon, the result of a complex, evolved and learned set of capacities. Philosopher of neuroscience Patricia Churchland argues in *Braintrust* that human moral behaviour emerges from the mechanisms that evolved to promote social interactions.

Churchland hypothesizes that ethics is "a four-dimensional scheme for social behaviour that is shaped by interlocking brain processes". She identifies these dimensions as: caring, rooted in a biological attachment to kin and kith; the ability to recognize others' psychological states; problem-solving in a social context; and social learning. Different brain networks contribute to each. And even simple moral decisions, such as whether to steal another's possession, involve at least a few of these dimensions.

Churchland describes the biological mechanisms that serve these functions in an accessible way, and makes a case that they are fundamental to our moral practices. Although there is no clear argument that these dimensions are distinct, interlocking or sufficient to account for all aspects of ethical thought, she succeeds in painting a picture of human morality as emerging from the social interactions of intelligent animals primed for cooperative, context-sensitive behaviour.

> *"It is still open to us to question what values are the best to espouse."*

The tour begins with the biology of "the most elemental" values: the neural hardware that grounds self-caring and well-being. Churchland's plausible claim is that our ability to value at all is rooted in the basic neural mechanisms that evolved to maintain human viability and reproductive success. These include neural pathways mediating negative and positive reinforcement learning, which underlie the experiences of pleasure, pain and fear; and hormones such as oxytocin and vasopressin, which figure in social interactions, pair bonding and care for offspring.

Thus, Churchland argues, human morality is formed on a neurobiological scaffold that evolved long ago, originally to promote care of self and offspring, and was later co-opted to foster more sophisticated cooperative interactions. She notes fascinating data that suggest that high-level social phenomena such as trust, generosity and punishment are influenced by levels of oxytocin and vasopressin.

Despite her naturalism, Churchland does not shy away from criticizing naturalistic views that she deems to be based on inadequate evidence. She argues that it is misleading to conceive of morality as innate, genetically specified or associated with one module in the brain. Although many social traits are genetically influenced, few can be linked to single genes. Instead, moral behaviour emerges as a solution to complex social problems.

She argues effectively against the popular view that humans share a "moral grammar" or a universal set of moral intuitions, but less convincingly against the position that human morality is built on a set of shared, emotion-involving sensitivities that become articulated differently in different cultures and environments. Churchland offers penetrating criticisms of arguments for the role of mirror neurons in theory of mind and empathy. These neurons respond when a monkey performs a certain goal-directed action, or sees that action being performed. Although mirror neurons are intriguing, they have been invoked to explain a wide variety of aspects of social neuroscience, often without adequate empirical or logical justification. Churchland's discussion puts these areas of research prone to over-interpretation into much-needed perspective.

Churchland demonstrates that there is substantial evidence that biology promotes social behaviour. But is all such behaviour necessarily good? Any behaviour, even free-riding, cheating and harming, must be supported by biology. How should we decide what is right? Many philosophers appeal to moral rules to address this question; Churchland sees no need, and criticizes those moral systems that take rules as foundational.

Churchland has Aristotelian sensibilities. The Aristotelian idea that morality depends on the inculcation of good behavioural habits that promote human flourishing fits well with the author's
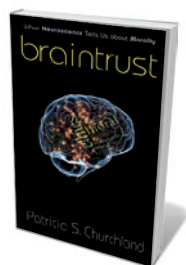
Braintrust: What Neuroscience Tells Us About Morality
PATRICIA S. CHURCHLAND
Princeton Univ. Press: 2011. 288 pp. $24.95

evolutionary approach. She is less charitable to some other moral philosophers, such as Immanuel Kant, who grounded morality in the operation of reason. Although Churchland's argument that morality has biological roots demonstrates that our moral behaviour is not grounded in reason alone, her quick dismissal of a central role for deliberation in moral cognition potentially deprives us of valuable insights about the nature of human morality. For example, it may be the ability to step back from our biological inclinations and deliberate abstractly that defines human prosocial behaviour as uniquely moral.

So how do we ascertain which of the competing biological drives, urges and intuitions should govern our behaviour? This brings us to the descriptive–normative distinction, the famously unbridgeable gulf between "is" and "ought". Churchland tries to argue that this gulf is more apparent than real. But in my view, she goes too far in denying the legitimacy of this question. I am sympathetic to her position that knowing how things are (facts) can influence deliberation about what should be done (values). However, it is still open to us to question what values are the best to espouse, or whether human flourishing is the only good to which we ought to aim.

This perplexing question of moral authority seems obvious to people of faith. In the final chapter, Churchland faces head-on the challenge from religion. She does an excellent job of pointing out the inconsistencies and dangers in taking religious authority as grounds for moral authority. But where that leaves us remains unclear. Are we realists, relativists or nihilists? Churchland's biology-centred picture leaves open significant questions about the good and the right that will require philosophical, not biological, answers.

In my view, by illuminating the biological foundations on which caring, cooperation and social understanding are based, and by arguing against simplistic views about innateness and divine ordination, Churchland has delineated the conceptual space still to be navigated concerning which actions are morally right, how we come to those decisions, and how we justify them. *Braintrust* covers what science tells us about morality. But that is only one side of the story. ∎

Adina L. Roskies *is an associate professor in the Department of Philosophy, Dartmouth College, New Hampshire 03755, USA.*
e-mail: adina.roskies@dartmouth.edu

Raymond McCauley tests dietary supplements on himself in his home lab.

BIOTECHNOLOGY

DIY biology

Bart Penders relishes an account of 'biohackers' who experiment beyond the confines of the lab.

In laboratories across the world, people with multiple degrees work at hygienic benches kitted out with globally recognized brands of scientific equipment. It is there, we assume, that the knowledge of how life works is assembled.

In his book *Biopunk*, science journalist Marcus Wohlsen unpeels a parallel universe of amateur knowledge-gathering — that of do-it-yourself biology, also known as biohacking. Wohlsen introduces us to a subculture of bioenthusiasts in the United States, Europe and Asia who tinker with life in their kitchens and garages. DIY biology may be less sterile and less well-equipped than that done in your average lab, but it is functional and fascinating, as his engaging account of biohackers shows.

Whether or not they have official qualifications, DIY biologists — or biopunks, as some of them dub themselves — are driven to study the stuff of life. They make up a colourful crew. Kay Aull, for instance, is a Boston-based lab technician and graduate of the Massachusetts Institute of Technology who designed and built a genetic test for the iron-overload disorder haemochromatosis in her apartment, using vintage equipment. Texan bioinformatician Raymond McCauley trawled through his own genome sequence in his home lab, even carrying out a small clinical trial, to test whether he can prevent macular degeneration through diet.

Biopunk: DIY Scientists Hack the Software of Life
MARCUS WOHLSEN
Current: 2011.
256 pp. $25.95

Wohlsen reminds us that professional scientists have never had a monopoly on knowledge-making. These days, professionals might be at the forefront initially, but often a science or technology develops in interesting ways only when it goes beyond the confines of a laboratory. Aull and McCauley's audacious experimentation exemplifies how shifts in the availability of data, reagents, technology and knowledge have altered how society interacts with the biological. Engaging with the genetic code is no longer just for the biotechnological elite.

That does not mean that DIY biology is an innocent pastime. A lot can be at stake. While experimenting, Aull and McCauley put their bodies on the line: blood, tissue and code. There is also a political agenda. As Wohlsen says, "Biopunks want to see whether the wall around the fortress of Big Science is really as high as it seems." The biopunk manifesto, declared by biohacker Meredith Patterson last year at the Outlaw Biology symposium at the University of California, Los Angeles, is a call for the democratization of science and a vow to put the tools of science into the hands of all (see go.nature.com/nyipbu).

But the drive has hit hurdles. The existing system of institution-based inquiry and innovation confines the tools and right to do biology to the elite, and thus resists democratization. Intellectual-property rights and restrictions on the availability of biological materials can prove problematic for home-based research.

But necessity forces invention and solidarity, Wohlsen explains. Biohackers are drawn towards open-source options, and make their research data and protocols freely available. Even so, their working conditions place limits on what DIY biologists can use: for instance, some dumped the preferred lab organism *Escherichia coli* in favour of *Lactobacillus* because it can endure garage conditions much better. As a result, Wohlsen writes, "the raw materials of biotechnology are always just a supermarket away", citing Patterson's use of spoilt milk and yoghurt to get bacterial cultures up and running.

Wohlsen urges us to reflect on how we organize science and how the power to define the biological is distributed. And he highlights the intrinsic messiness of innovation, using historical examples of experiments by the gentlemen scientists of yesteryear that we might now consider to be ethically dubious. Wohlsen reminds us that every encounter with the unknown is inherently uncertain, and potentially brings new ethical challenges. Some authorities might also suspect the motives of people who mess about with Petri dishes, flasks, chemicals and microorganisms in their basements, leading to allegations of bioterrorism. The position of DIY biology in our society has yet to be established.

In my view, no clear line can or should be drawn between institutional life sciences and DIY biology, and the dichotomy between expert and layperson does not stand firm. Most biopunks, including Aull and McCauley, have jobs at universities or in biotech companies. Ultimately, biohacking is more concerned with getting biology out of the lab and into the streets than with making major scientific breakthroughs. The objective, after all, is the democratization of science. French sociologist of science Michel Callon has called it "research-in-the-wild", in which the professional and the amateur increasingly overlap.

Biopunk carries a convincing message that slowly, steadily, we could all become authors of a ubiquitous and democratic biology. ∎

Bart Penders *is a postdoctoral researcher in the Department of Health, Ethics and Society at Maastricht University, the Netherlands, and a senior researcher at the Institute for Science, Innovation and Society at Radboud University Nijmegen, the Netherlands.*
e-mail: b.penders@maastrichtuniversity.nl

David Meyer, one of three actors — including a woman — who play different facets of Isaac Newton in a play about his life.

THEATRE

Newton's rainbow

A biographical play reveals the odd character of the father of gravity, finds **Philip Ball**.

Isaac Newton perplexes and fascinates both as a transitional figure in the history of science and because he was a very odd man. The difficulty has been in distinguishing the two. Historians still struggle to dispel portrayals of him as a man torn between science and religion, or flitting from mathematical physics to superstitious alchemy. It was not, however, the coexistence of these things in Newton's agenda that made him odd — that was not unusual in his time — but the way he lived, isolated from intimate relationships, sensitive to every slight, vain yet so indifferent to posterity that he could barely be persuaded to write the *Principia*.

All this makes him attractive and challenging to biographers, among them the leading science historians Richard Westfall and A. Rupert Hall, and science writer James Gleick. It has also inspired some more inventive explorations of his life story, the latest of which is *Let Newton Be!* This play by Craig Baxter was commissioned by the Faraday Institute for Science and Religion at the University of Cambridge, UK, after the institute's director saw Baxter's 2007 play about Charles Darwin. The production has benefited from the input of historian Rob Iliffe, head of the Newton Project to place the scientist's works online, and the astrophysicist John Barrow, among others.

To conjure up the mercurial mathematician, Baxter uses mostly Newton's own words and those of some of his contemporaries, such as his rival and critic Gottfried Leibniz. Newton is the only character in the piece, apart from brief appearances by the likes of Leibniz and Edmond Halley, and is played by three actors, one a woman. It sounds like a gimmick, but it is actually a clever device that allows us to see different facets of the man.

The play's structure is largely chronological. We see Newton as a boy in the family home at Woolsthorpe in Lincolnshire; as an undergraduate at Trinity College, Cambridge; then as Lucasian professor of mathematics, a post to which he was appointed in 1669 aged 27. We see him take his retractable telescope to the Royal Society and then, stung by what he perceives as the antagonism of the London virtuosi, retreat into religious exegesis until Halley cajoles him into writing down his proof of elliptical planetary orbits, a treatise that expands into the *Principia*. Fêted and now somewhat pompous, he becomes warden of the Royal Mint and president of the Royal Society.

The original material is well used. There is a reconstruction of Newton's famous prism experiment, or roughly so — this *experimentum crucis* of around 1666, when he reconstituted white light from the spectrum, is notoriously difficult to reproduce. We are introduced to Newton's obsessive, sometimes surreal lists of the sins he committed: "I lied about a Louse." The only time his words turn into a lecture is intentional, when we witness one of the optics lectures that Newton was obliged to give as Lucasian professor, at which he

proves to be hilariously inept.

The play delivers an impressive quantity of Newton's thought. In particular, it emphasizes just how much of his work was religious. Newton considered this to be his central mission, with the seminal scientific works on light, motion and gravity almost being tossed off before breakfast. The idea that exploring the natural world allowed a deeper appreciation of God's wisdom and power was the position held by most seventeenth-century scientists, and their defence against accusations of materialistic atheism. Newton was anything but a materialist. That he held gravity to be an occult force acting at a distance was precisely what Leibniz considered wrong with his theory, whereas for Newton, this force was actively sustained by God.

But I am not sure how much of this material would be comprehensible to anyone coming to Newton anew. It is characteristic of the play's intelligence that we do not get any nonsense with falling apples. But neither are we told, say, what distinguished Newton's ideas on gravity from the many that went before, especially René Descartes' vortices and the belief that it is a form of magnetism, both views that Newton shared at some point. And the play lacks a narrative drive — there is no tension, nothing to be resolved, for in the end it is still a kind of biography. But that was its brief, and it is probably a more enjoyable hour and a half with Newton than anyone ever had in his lifetime. ∎

Let Newton Be!
WRITTEN BY CRAIG BAXTER; DIRECTED BY PATRICK MORRIS.
*London, 28 March.
Now touring North America until 30 April.*

Philip Ball *is a writer based in London. His latest book is* Unnatural.

CORRESPONDENCE

FDA not NIH can speed new drugs

Finding ways to accelerate the development of new pharmaceuticals is a worthy goal, but delegating the task to the US National Institutes of Health (NIH) is a bad idea (*Nature* **471,** 135; 2011). Change is needed at the US Food and Drug Administration (FDA), not the NIH.

The NIH has long excelled at funding and conducting basic research, but it should leave the development of commercial products to the private sector. Expenditures on commercial enterprises are likely to reduce the funding available for basic research.

Government bureaucrats are rarely leaders of technological innovation. They are not qualified to act as venture capitalists in choosing the most promising and deserving commercial product to fund and develop.

NIH director Francis Collins is right to be frustrated at how many basic science discoveries with therapeutic potential are waiting for follow-up by the pharmaceutical industry (G. Harris, *The New York Times*, 22 January 2011). But his colleagues at the FDA bear much of the responsibility for drug companies' reluctance or inability to take these forward.

Drug approvals have been dismal at a time when drug development should have been spurred by spending and new technologies. Expenditure on research and development by US-based drug companies almost quadrupled to more than US$65 billion between 1995 and 2009.

Fred Hassan, former chief executive of drug company Schering-Plough, asked "What will it take to get new drugs approved? The point is, we don't know." Kenneth Kaitin, director of the Tufts Center for the Study of Drug Development, blamed the FDA's 'obstructionist' culture for its reputation as "an agency that is supposed to keep unsafe drugs off the market, not to speed access to lifesaving drugs" (*The Wall Street Journal*, 30 June 2008).

Henry I. Miller *Hoover Institution, Stanford University, USA.*
henry.miller@stanford.edu

Libyan uprising may boost bluefin tuna

The unrest in the Middle East could provide an unexpected shield for endangered stocks of bluefin tuna (*Thunnus thynnus*) around the Libyan coast.

From 2002, fishing operations off Libya have been highly lucrative. Vessels often catch tuna worth €500,000 (about US$712,000) in a single haul. The Libyan government banned foreign tuna fishing in its exclusive economic zone from 2004. But the French tuna-fishing fleet negotiated Libyan licences for some of its vessels, and immediately resumed fishing for Mediterranean bluefins.

Then came the Libyan revolution and the military intervention by France and its allies. The French tuna-fishing fleet is likely to lose its access to Libyan waters for the foreseeable future. This is good news for tuna. Like North Atlantic fish stocks during the First and Second World Wars, Mediterranean bluefins are likely to escape trawlers as long as military operations are under way. Such a fishing curfew could be more effective than decades of recommendations by the International Commission for the Conservation of Atlantic Tunas. A similar situation effectively stopped tuna fishing in the western Indian Ocean from 2007 — as a result of Somalian piracy.

David Grémillet *Centre for Functional and Evolutionary Ecology, CNRS, UMR5175, Montpellier, France; and PFIAO, DST/NRF Centre of Excellence, University of Cape Town, South Africa. david.gremillet@cefe.cnrs.fr*
Amélie Lescroël *University of Rennes and the National Museum of Natural History, UMR7204, Rennes, France.*

Predicting the threat of nuclear disasters

The failures at Japan's Fukushima Daiichi nuclear power plant after the tsunami on 11 March have led people to question the risks associated with nuclear power stations. But disasters on this scale can be considered to be very unlikely if it is assumed, like natural catastrophes, that the more energy that is required to cause an event, the less likely it is to happen.

In the biosphere, for example, extinctions of single species or communities of organisms occur much more frequently than mass extinctions. Such natural events show a logarithmic scale of frequency distribution when plotted against energy input.

Soon after the Chernobyl disaster, Kenneth Hsü suggested that man-made accidents might show a similar frequency–magnitude relationship (*Nature* **328,** 22; 1987). He proposed that financial costs could be used as a measure of magnitude, because every small accident has a monetary equivalent — for example, in lost electricity-supply payments and insurance-claim payouts. Everything from the smallest mishaps (such as temporary shutdowns) to the largest catastrophes could thus provide a wide range of data points, putting statistical assessment of nuclear-accident probabilities on a firmer footing than, say, radiation-leak measurements (*Nature* **335,** 391; 1988). It may be time to take a fresh look at Hsü's proposals for gauging the safety of nuclear power plants.

A. M. Celâl Şengör *Istanbul Technical University, Turkey.*
sengor@itu.edu.tr

Regional drought has a global impact

In today's interconnected world, regional droughts can have a global impact on food supplies and undermine the economics and stability of governments.

For example, droughts last year in Russia and Ukraine reduced the wheat harvest by 32.7% and 19.3%, respectively, severely diminishing the worldwide wheat supply. China, the world's largest wheat producer and consumer, is being forced to import grain after a continuing drought in its northern growing region (*Nature* **470,** 307; 2011).

The fall in wheat production has contributed to a sharp rise in global prices — by February 2011, these had more than doubled in 8 months. The economies of wheat-importing countries in the Middle East and North Africa are particularly hard hit. This is especially true of Egypt, the world's largest wheat importer, and Tunisia and Libya, where the steep rise in the cost of food is fuelling political and economic dissatisfaction.

Troy Sternberg *University of Oxford, UK.*
troy.sternberg@geog.ox.ac.uk

Simon van der Meer
(1925–2011)

Engineer whose invention enabled the discovery of the W and Z particles.

Simon van der Meer was one of the handful of truly exceptional people who contributed to making the European Organization for Nuclear Research (CERN) the world's premier laboratory in elementary particle physics. His invention of stochastic cooling offered a means to nudge protons and their antimatter equivalents into tightly focused beams, paving the way for the discovery of the W and Z particles — two fundamental constituents of matter.

Van der Meer, who died on 4 March, studied technical physics at the University of Technology in Delft, the Netherlands, from 1945, at a time when Dutch universities were still recovering after being closed under the German occupation. He eventually obtained his engineering degree in 1952. Following a short period in Eindhoven at the Philips Research Laboratory, he was among the first researchers to move to the recently founded CERN near Geneva, Switzerland. He arrived at the age of 31 in 1956, a year before the laboratory's first accelerator was built, and worked there for the rest of his life.

In the early 1970s, circumstantial evidence for the electroweak model was accumulating. This model unified electromagnetism and the weak nuclear force, two of the four fundamental forces of nature (gravity and the strong nuclear force being the other two) in a single mathematical framework. But no one would be convinced until physicists had actually observed the W and Z bosons — particles that supposedly carried the weak force.

Accelerators at the time could smash particles into a stationary object. But producing the heavy W and Z bosons would require much more energy — specifically, getting protons and antiprotons to collide head on in tightly packed beams.

Among van der Meer's many original and highly practical ideas, stochastic cooling was the most important. This was, in essence, a subtle and practical realization of the 'Maxwell's demon' thought experiment, which demonstrates how, under certain circumstances, the second law of thermodynamics (which states that differences in temperature, pressure and so on equilibrate over time) can be circumvented.

It is universally accepted among physicists that any beam of elementary particles should obey the theorem of the French mathematician Joseph Liouville, and behave like an incompressible fluid. In other words, if the beam is squashed at one end, the distribution of particles will bulge out somewhere else along its length. Beams of fast-moving electrons and positrons, their antiparticles, disobey this rule because they emit a form of radiation; in essence, the beams 'cool' naturally. But in the case of the more massive particle beams, the Liouville theorem kicks in. This presents a significant obstacle to producing an orderly, narrow beam.

Van der Meer's idea was both simple and ingenious. The 'incompressible fluid' condition is strictly valid only when very large numbers of particles are present. When small numbers of particles circulate through an accelerator ring, small fluctuations from the average course taken by the particles are observed. In the 1970s, these fluctuations — called Schottky noise — could easily be detected using a device called a pick-up electrode. Because every particle in the beam has a specific individual energy, different amounts of fluctuation are detected each time the beam of particles courses through the ring.

Van der Meer's idea was to 'correct' the average position of the particles in a beam electronically with the help of a 'kicker': a magnetic device that continuously nudges the particles back on course. This very subtle form of compression of the beam does not violate the Liouville theorem. Indeed, through the synchronized use of a pick-up and a kicker, the very tiny elementary volumes around which a particle is present are simply moved away from those around which there are no particles. The density of particles seems to increase simply because the voids have been pushed to the side.

The approach was originally applied to beams of protons, where the stochastic cooling effect was visible but very small because the number of particles in the beams was very large. In the late 1970s, David Cline, Peter McIntyre and I used a magnetic field to steer head-on protons and antiprotons, colliding them many times every second. Provided that we ramped up the density of the antiprotons as much as a billion times, van der Meer's cooling process became prominent.

Stochastic antiproton cooling was an essential ingredient for the realization of the CERN ppbar collider and for the subsequent discovery of the W and Z particles — for which van der Meer and I shared a Nobel prize in 1984. The same technology has been key at the Tevatron collider at Fermilab in Batavia, Illinois, where the top quark, the last and heaviest of all elementary quarks, was discovered in 1995. Indeed, during the past 25 years or so, stochastic cooling has been a major tool for all the significant discoveries made from high-energy collisions.

It is very unusual for a clever idea of an engineer to deeply modify fundamental concepts in physics. Simon himself noted that his lack of formal training in physics (he never obtained a PhD) probably turned out to be an asset. Not knowing the rules of the game meant he came up with an unthinkable idea.

With Simon's departure, we will miss not only a remarkable, inventive mind and a giant in our field, but also a very kind and universally respected person. His friends used to say that he would never use two words where one would suffice, but that his one word would invariably be the right one. ∎

Carlo Rubbia is at CERN, Geneva 23, Switzerland.
e-mail: carlo.rubbia@cern.ch

NEWS & VIEWS

Spin drag in a perfect fluid

Beautiful experiments in an ultracold, strongly interacting atomic Fermi gas reveal that, unexpectedly, unpaired atomic spins flow in the system with the maximum resistance permitted by the laws of quantum mechanics. SEE LETTER P.201

JOHN E. THOMAS

The transport of mass and spin in strongly interacting quantum fluids is of interest in fields ranging from high-temperature superconductivity to nuclear physics. Counterintuitively, strongly interacting systems, at temperatures at which they are not superfluids, can be nearly perfect fluids, with almost frictionless mass flow[1]. However, on page 201 of this issue, Sommer *et al.*[2] demonstrate that unpaired spins flow with the maximum possible resistance in such systems. An increased understanding of the relationship between the transport properties of mass and spin should shed new light on spin pairing and atomic collisions in strongly interacting fluids.

Spin has an important role in quantum physics. In some ways, it is analogous to a toy spinning top rotating about its axis. In contrast to toy tops, however, quantum particles that have the smallest possible non-zero spin, ½, have just two states (spin orientations), pointing either 'up' or 'down', similar to an electron in an atom. Furthermore, two spin-½ particles in the same state (both up or both down) avoid one another and cannot collide — they are fermions, which obey Pauli's exclusion principle.

The 'digital' up–down property of spin-½ particles is exploited in the field of spintronics, in which up and down correspond to the binary 0 and 1 of a digital computer and information is carried by the spin rather than the charge of an electron. In the same way that an electric current arises from the difference between the flows of positive and negative charges, a spin current arises from the difference between the flows of up and down spins.

An ultracold atomic Fermi gas provides an ideal, precisely controllable model system for exploring mass and spin currents in a strongly interacting quantum fluid. Experiments usually use a gas of spin-½-up and spin-½-down atoms in an optical trap, which holds the atoms in place by means of light. An external magnetic field controls the interaction between spin-up and spin-down atoms, and by tuning the magnetic field to the correct value, opposite spins interact as strongly as permitted by the laws of quantum physics. Using the technique of evaporative cooling,
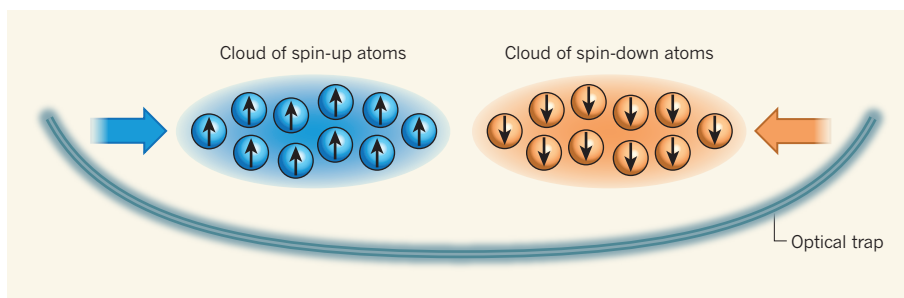
Figure 1 | Bouncing atomic clouds in an optical trap. Sommer *et al.*[2] used a magnetic-field gradient to separate clouds of spin-up and spin-down fermionic atoms in an optical trap. When the clouds were released, by turning off the magnetic-field gradient, they were pushed towards each other by the optical trap, which acts like a bowl. The strongly interacting clouds bounce off each other several times and then penetrate each other slowly, on a timescale of a second, demonstrating extremely high resistance to spin flow.

the gas temperature can be lowered into the quantum degenerate regime. In this regime, the atomic momentum p is so small that the de Broglie wavelength, $\lambda = h/p$ (where h is Planck's constant), is comparable to the interatom spacing and the Pauli exclusion principle comes into play.

Now, how do mass and spin flow in a strongly interacting Fermi gas? One might expect that a strongly interacting gas would be 'sticky', and would flow with maximum viscosity. In fact, recent measurements[3] show that the system's mass flow has an extremely low viscosity and, remarkably, the experiments reported by Sommer *et al.*[2] reveal the opposite for the spin current, where maximum resistance to spin flow is observed.

This apparent contradiction can be understood using a simple physical argument made by Sommer and colleagues. Shear viscosity determines how efficiently force is transmitted between two plates by a fluid lying between them. Moving the lower plate to the right causes some nearby atoms to acquire an average velocity to the right. These atoms can move upward to transmit force to the right through neighbouring sheets of atoms. When the atoms can move upward easily without colliding with each other (they have a large mean free path), a large force is transmitted to the upper plate, and so the viscosity is high. By contrast, for strongly interacting systems, the atoms travel only a very short distance before colliding (they have a very small mean free path), resulting in a very small transmitted

force and consequently a low viscosity.

The mean free path is minimized when the collision rate between atoms is highest. However, when the temperature of a Fermi gas is lowered into the degenerate regime, atoms have difficulty colliding with one another. In an ordinary Fermi 'liquid', the decreasing collision rate would increase the mean free path, causing the viscosity to increase as the temperature is lowered. Instead, the viscosity decreases with decreasing temperature[3]. A possible explanation is that strong pairing interactions between spin-up and spin-down atoms form spin-zero particles (bosons), which 'collide' at an enhanced rate, because bosons prefer to be near each other. This process may be responsible for the very low viscosity[4].

If pairing is responsible for the low viscosity observed in mass flow, what happens when individual spins move in a strongly interacting system? To investigate this question, Sommer and colleagues[2] prepared a unique system: two separated clouds of atomic fermions, one with spin up and the other with spin down, contained in a shallow optical trap (Fig. 1). When the clouds were released, by turning off the magnetic-field gradient that had initially kept them apart, they were pushed towards each other by the optical trap, which acts like a bowl. Remarkably, the clouds then bounced off one another. The authors suggest that the bouncing arises because strongly interacting clouds of opposite spin have a short mean free path when travelling forward into each other — the

spins in each cloud are then more likely to collide, reverse their velocity and travel freely backward, because in the backward direction all atoms are in the same spin state and are prevented from colliding by the Pauli exclusion principle.

After the bouncing subsided, the behaviour of the clouds was much like that of two masses connected by an oil dashpot that heavily damps their relative motion. The measured damping exhibited a maximum at a certain temperature and decreased as the temperature

was further reduced, suggesting that the collision rate was suppressed at low temperature, as expected for a Fermi liquid, and that the resistance to spin flow decreased. From these measurements, Sommer et al.[2] conclude that, for spin currents, pair formation cannot produce bosons to enhance the collision rate: spin-up and spin-down atoms must move in opposite directions to produce a net spin current, and so pair formation cannot occur as it does for mass flow. Thus, developing a detailed understanding of transport in Fermi

gases presents further challenges for the theory of strongly correlated many-body systems. ■

John E. Thomas is in the Department of Physics, Duke University, Durham, North Carolina 27708, USA.
e-mail: jet@phy.duke.edu

1. Thomas, J. E. Phys. Today 63(5), 34–37 (2010).
2. Sommer, A., Ku, M., Roati, G. & Zwierlein, M. W. Nature 472, 201–204 (2011).
3. Cao, C. et al. Science 331, 58–61 (2011).
4. Bruun, G. M. & Smith, H. Phys. Rev. A 75, 043612 (2007).

NEUROSCIENCE

Channelopathies have many faces

A sodium channel known for its role in the perception of pain also seems to be necessary for olfaction. The multiple roles of this channel and the diverse effects of its mutations raise intriguing questions. SEE ARTICLE P.186

STEPHEN G. WAXMAN

Generations of scientists were taught that 'the' voltage-gated sodium channel serves as a molecular battery, producing electrical impulses in nerve and muscle cells. However, we now know that mammals, including humans, have nine sodium-channel isoforms ($Na_V1.1$–$Na_V1.9$), encoded by different genes. Of these, $Na_V1.7$ has been the focus of much recent attention, because it is a major contributor to pain perception and so a therapeutic target for pain management[1]. But, as Weiss et al.[2] demonstrate on page 186 of this issue, it is becoming increasingly clear that $Na_V1.7$ also has other neurosensory functions.

$Na_V1.7$ is preferentially expressed in the peripheral nervous system — in dorsal root ganglion (DRG) neurons and sympathetic ganglion neurons[3], including the peripheral axonal termini of neurons that perceive pain[4]. The first insight into the role of this channel in human neurosensory processes came with the discovery that mutations in $Na_V1.7$ cause human pain disorders.

Ion-channel mutations can enhance (gain-of-function) or attenuate (loss-of-function) channel activity. For instance, gain-of-function mutations in $Na_V1.7$, which make it easier to activate this channel and so increase the excitability of pain-signalling DRG neurons, cause inherited erythromelalgia (IEM) — a condition in which, on exposure to mild warmth, patients experience severe burning pain and redness of the skin on the limbs[5]. Another group of $Na_V1.7$ gain-of-function mutations that interfere with inactivation (transitory silencing of the channel after it has been

activated) cause paroxysmal extreme pain disorder (PEPD)[6]. Patients with PEPD are affected by episodes of pain in the lower body, eyes and jaw.

By contrast, loss-of-function mutations in $Na_V1.7$ — including nonsense, frame-shift and splicing mutations — prevent the production

of functional $Na_V1.7$ channels and cause a disorder called channelopathy-associated insensitivity to pain (CIP)[7]. Patients with CIP do not experience pain even when confronted with extremely painful stimuli: bone fractures, dental extractions and burns cause them no pain.

A hint that $Na_V1.7$ might participate in olfaction came from observations[8] that mice lacking $Na_V1.7$ die soon after birth; this was proposed to be due to failure to feed, possibly because the animals could not respond to olfactory cues. Anosmia, or inability to sense smell, has more recently been reported[9] in patients with CIP. Weiss et al.[2] now show that $Na_V1.7$ plays a crucial, non-redundant part in odour perception. They demonstrate that $Na_V1.7$ is present in axons of human olfactory sensory neurons, and show in mice that it is essential for the initiation of synaptic transmission from olfactory sensory neurons to
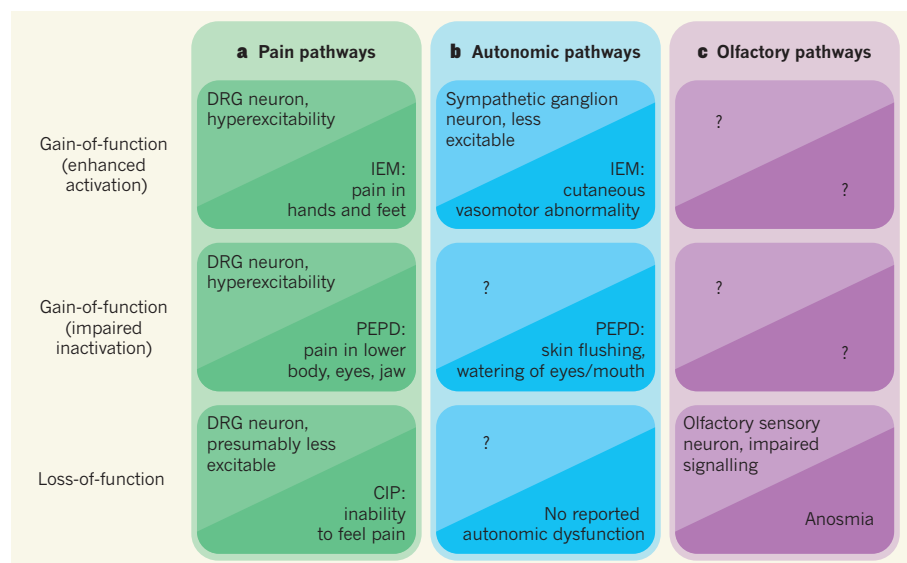
Figure 1 | Multiple effects of mutations in the $Na_V1.7$ ion channel. a, Gain-of-function mutations in $Na_V1.7$ increase the excitability of dorsal root ganglion (DRG) neurons, thereby producing pain in inherited erythromelalgia (IEM) and paroxysmal extreme pain disorder (PEPD). Loss-of-function mutations of this channel decrease excitability in DRG neurons, and produce channelopathy-associated insensitivity to pain (CIP). **b,** Gain-of-function mutations of $Na_V1.7$ reduce the excitability of sympathetic ganglion neurons in IEM, thereby interfering with vasomotor (blood-vessel calibre) control in the skin. They may also perturb the firing of sympathetic ganglion neurons in PEPD, in which skin flushing, watering of the eyes and mouth, and slowed heart rate accompany pain attacks. **c,** Weiss et al.[2] show that, in the olfactory system, loss-of-function mutations in $Na_V1.7$ interfere with signalling from olfactory sensory neurons to upstream nerve cells, causing anosmia, the inability to smell.

higher-order neurons that project to the brain.

The finding that Na$_V$1.7 has a pivotal role in olfactory sensory processing presents an interesting parallel to earlier observations that this channel plays a crucial part in pain signalling[1]. A strategy that is currently being explored for treating pain is specific blockade of the Na$_V$1.7 channel. Whether this approach will be complicated by impaired ability to smell and, if so, whether this will be clinically significant, remain to be determined.

Another question is whether the Na$_V$1.7 gain-of-function mutations that cause IEM and PEPD affect olfactory processing. Na$_V$1.7 mutations responsible for IEM produce hyperexcitability in DRG neurons, which causes pain, but make sympathetic ganglion neurons less excitable, thereby interfering with the regulation of blood flow to the skin[10]. This complex relationship between genes and their associated traits demonstrates that a single mutation can have very different effects on cellular function when expressed against different cell backgrounds (Fig. 1).

IEM mutations have multidirectional effects because, in addition to enhancing activation, they can depolarize the neuronal membrane[10]. DRG neurons also contain Na$_V$1.8 sodium channels which, unlike other sodium-channel isoforms, are not inactivated by this level of depolarization[11]. The depolarization produced by IEM mutations brings DRG neurons closer to the threshold for activation of the Na$_V$1.8 channels, thereby contributing to hyperexcitability of these cells[12]. By contrast, this depolarization inactivates and effectively silences almost all sodium channels present in sympathetic ganglion neurons — these cells lack Na$_V$1.8 — making it harder for them to fire[10]. These data hint that the effects of IEM and PEPD mutations on olfactory sensory neurons probably depend on the as yet unknown complement of other ion channels in these cells.

Further complicating the story are compensatory and post-translational changes, which may modulate the effects of mutated ion channels. Na$_V$1.7 is normally present in sympathetic ganglion neurons as well as DRG neurons. Yet, despite the lack of functional Na$_V$1.7 channels in CIP, dysfunction of the autonomic nervous system, which includes sympathetic ganglion neurons, is absent, or relatively subtle, in this disorder[7,9]. It is not clear whether the relative absence of autonomic dysfunction in CIP is due to the presence of redundant sodium-channel isoforms that can take over the duties of Na$_V$1.7 in sympathetic neurons; to compensatory increases in the expression of other sodium-channel isoforms in these cells; or to increased sensitivity of upstream neurons (denervation hypersensitivity).

Similarly, we do not understand why gain-of-function mutations that enhance the activation of Na$_V$1.7 channels produce pain primarily in the hands and feet (IEM)[5], whereas those that affect channel inactivation produce

lower-body, eye and jaw pain (PEPD)[6]. Multiple binding partners, including protein kinases[13], modulate Na$_V$1.7, and many other ion channels, but the clinical consequences of this modulation remain unknown.

Taken together, these observations suggest that a combination of factors — including cell-background-dependent and activity-dependent events, epigenetic factors and environmental influences — regulates the activity of normal ion channels, and can modulate the effects of ion-channel mutations even in disorders such as IEM, PEPD and CIP, which are monogenic in the sense that they are produced by a single gene mutation. The full span of these factors and their effects on normal ion channels, as well as the full extent of the ever-expanding range of cellular and clinical abnormalities caused by ion-channel mutations, will undoubtedly become clearer in the next few years. ∎

Stephen G. Waxman *is in the Department of Neurology and the Center for Neuroscience and Regeneration Research, Yale University School of Medicine, New Haven, Connecticut 06510, USA. He is also at the Veterans Affairs Connecticut Healthcare System, West Haven.*
e-mail: stephen.waxman@yale.edu

1. Waxman, S. G. *Nature* **444,** 831–832 (2006).
2. Weiss, J. *et al. Nature* **472,** 186–190 (2011).
3. Toledo-Aral, J. J. *et al. Proc. Natl Acad. Sci. USA* **94,** 1527–1532 (1997).
4. Persson, A.-K. *et al. Mol. Pain* **6,** 84 (2010).
5. Dib-Hajj, S. D. *et al. Brain* **128,** 1847–1854 (2005).
6. Fertleman, C. R. *et al. Neuron* **52,** 767–774 (2006).
7. Cox, J. J. *et al. Nature* **444,** 894–898 (2006).
8. Nassar, M. A. *et al. Proc. Natl Acad. Sci. USA* **101,** 12706–12711 (2004).
9. Goldberg, Y. P. *et al. Clin. Genet.* **71,** 311–319 (2007).
10. Rush, A. M. *et al. Proc. Natl Acad. Sci. USA* **103,** 8245–8250 (2006).
11. Akopian, A. N., Sivilotti, L. & Wood, J. N. *Nature* **379,** 257–262 (1996).
12. Harty, T. P. *et al. J. Neurosci.* **26,** 12566–12575 (2006).
13. Stamboulian, S. *et al. J. Neurosci.* **30,** 1637–1647 (2010).

MAMMALIAN EVOLUTION

A jaw–dropping ear

A fossil from the Early Cretaceous provides insight into the evolution of the hearing apparatus in mammals. Anchoring the eardrum was, it seems, an essential step in freeing the middle ear from the jaw. SEE ARTICLE P.181

ANNE WEIL

The 120-million-year-old lake sediments of the Jiufotang Formation of Liaoning, China, have yielded another scientific bounty: the near-complete skeleton of a fossil mammal, *Liaoconodon*. This fossil, described by Meng *et al.* (page 181 of this issue)[1], is notable in illuminating a key phase in the evolution of the mammalian ear.

The exceptional preservation of fossils in the Jiufotang Formation, which has offered up entire fish, pterosaurs and feathered dinosaurs, is no longer a surprise. But the cornucopia of knowledge provided by such detail seems endless. When the new specimen's small, flattened but complete skull was prepared, the palaeontologists found that the tiny bones of the middle ear were still in place, and still attached to the lower jaw[1]. In all extant mammals, they are separated from it. Their attachment in *Liaoconodon* might support the contention that the ear bones remained tenuously attached to the jaw higher in the evolutionary tree of mammals than some have supposed[2].

The middle ears of living mammals contain three small bones, the malleus, incus and stapes, which conduct vibrations from the eardrum to the inner ear. The eardrum itself stretches across a fourth small bone, the ectotympanic. The malleus and ectotympanic have long been understood to have been primitively

part of the lower jaw, or mandible; the joint between the malleus and incus is homologous to the jaw joint in reptiles. How, when and how many times these ossicles detached from the mandible during the course of mammalian evolution is a topic of some controversy[2].

Liaoconodon is a member of a lineage of mammals, the eutriconodonts, that is well-documented and diverse in the Liaoning beds, but has no living representatives. Eutriconodonts are unusual among vertebrates in having had an ossified Meckel's cartilage[3,4]. In extant mammals, Meckel's cartilage is a rod that condenses early in an individual's development and is later sheathed in dermal bone that will become the lower jaw. The cartilage itself is resorbed, leaving only its posterior end, which ossifies and becomes the posterior part of the malleus. In eutriconodonts, instead of degenerating, Meckel's cartilage separated from the part of the malleus it formed and then turned to bone, persisting in adult individuals. It is this element that the ear ossicles are contacting.

The exquisite preservation of *Liaoconodon* allowed Meng *et al.*[1] to identify the suture between the anterior and posterior parts of the malleus; it is the anterior process, derived not from cartilage but from a dermal bone of the jaw, that broadly contacts the ossified Meckel's cartilage and wraps part-way around it. The ectotympanic mostly contacts the
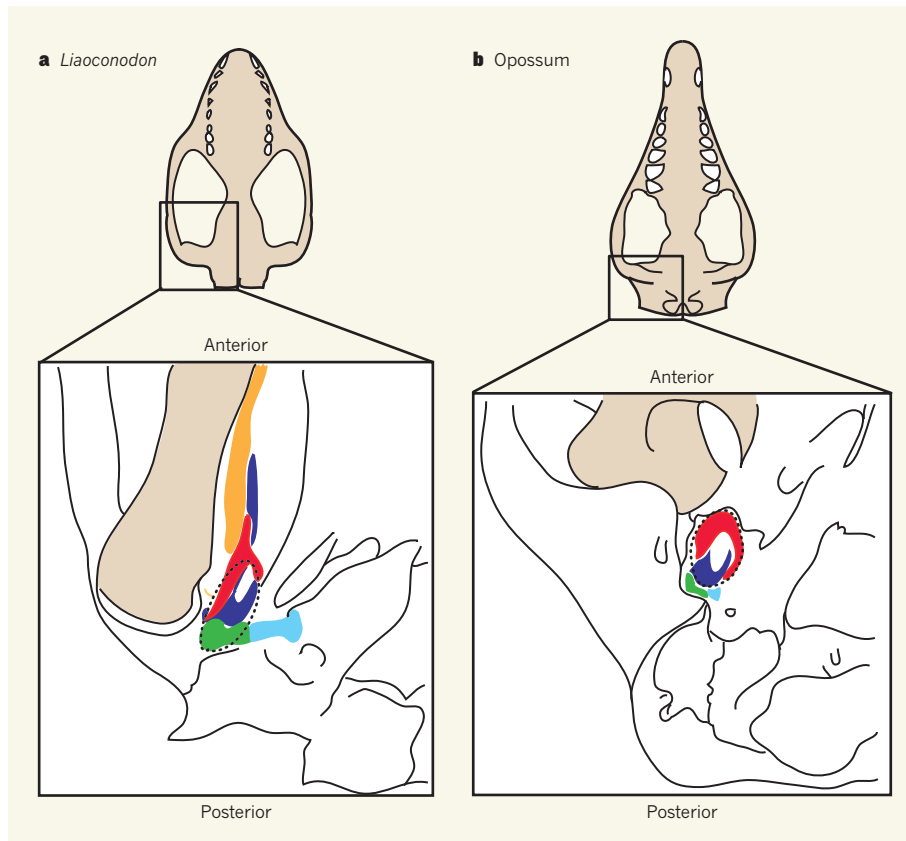
Figure 1 | Middle ears and eardrums in ventral view.   a, In the transitional ear of *Liaoconodon*, as inferred by Meng *et al.*[1], the eardrum (dotted line) would have been stretched between the ectotympanic bone (red) and the skull. Tension on the eardrum was maintained by the buttressing of the ectotympanic on the anterior process of the malleus (dark blue) and the ossified Meckel's cartilage (orange). **b**, In the ear of an extant mammal, the opossum *Didelphis*, the ectotympanic is fixed to the skull and the eardrum is almost entirely attached to it. The other ossicles depicted are the incus (green) and the stapes (light blue). The bones depicted in light brown are the posterior of the jaw showing the joint with the skull. (Insets modified from Fig. 3e and g of ref. 1.)

malleus. An adult euriconodont thus exhibits a morphology very similar to that of the fetus of an extant mammal at a certain stage of development. Contact of the ossified cartilage with a separated malleus had been inferred[3], but *Liaoconodon* offers our first opportunity to see the ossified cartilage in contact with the bones of the middle ear in an adult.

The transitional evolutionary state exemplified by *Liaoconodon* had to function in the adult animal. What role would a persistent Meckel's cartilage have played in hearing? *Liaoconodon*'s ectotympanic, free of the skull and describing approximately 90° of a rough circle, could only have supported the eardrum's anterior margin (Fig. 1a). In living mammals such as the opossum *Didelphis*, the ectotympanic generally describes an arc of 270° or more, supports almost the entire membrane, and is attached to the skull (Fig. 1b). Meng *et al.* argue that, to maintain the membrane at constant tension, the posterior part of *Liaoconodon*'s eardrum would have had to have been anchored to the skull and the ectotympanic would have needed stabilization provided by the malleus and ossified Meckel's cartilage. Functionally speaking, the ossified

Meckel's cartilage is more a part of the ear than of the jaw it is still attached to.

This functional explanation adds a dimension to discussion of how many times full separation of the middle-ear bones from the jaw evolved. There are two sides to this argument. One possibility is that the fully separated ear evolved once, was present in the common ancestor of mammals, and that euriconodonts and a group called the spalacotheroids, which also have an ossified Meckel's cartilage[4,5], represent evolutionary reversals[3,5]. Such reversals are, in theory, not difficult to accomplish; it would require a slight alteration in developmental timing[5].

The other possibility is that the common ancestor of mammals retained at least Meckel's cartilage, if not an even more primitive middle ear, as an adult, and that the fully separated ear evolved in three lineages within mammals: those leading to monotremes, such as the platypus; to the extinct multituberculates; and to the therians, which include extant marsupials and placentals. In this scenario, which Meng *et al.*[1] support, euriconodonts retain a transitional ear, more primitive than any found in living mammals.

Most early mammals are known only from fragments, and the persistence of Meckel's cartilage, or even the connection of the ear bones to the principal part of the mandible, can only be inferred from the grooves they nestled in along the internal side of the jaw. The second scenario has supporting points, however. First, a structure interpreted as a groove for Meckel's cartilage[4] is present in many extinct animals considered to be mammals[1,6]. Furthermore, phylogenetic models place monotremes and multituberculates as close relatives to animals not believed to have had fully separated middle-ear ossicles[3]. Understanding that Meckel's cartilage had a specific function crucial to hearing makes it even more plausible that the transitional state seen in *Liaoconodon* could have preceded evolution of the modern ear in multiple mammalian lineages. ∎

Center for Health Sciences, Tulsa, Oklahoma 74107, USA.
e-mail: anne.weil@okstate.edu

Anne Weil *is in the Department of Anatomy and Cell Biology, Oklahoma State University*

1. Meng, J., Wang, Y. & Li, C. *Nature* **472,** 181–185 (2011).
2. Martin, T. & Ruf, I. *Science* **326,** 243–244 (2009).
3. Luo, Z.-X., Chen, P., Li, G. & Chen, M. *Nature* **446,** 288–293 (2007).
4. Meng, J., Hu, Y., Wang, Y. & Li, C. *Zool. J. Linn. Soc.* **138,** 431–448 (2003).
5. Ji, Q., Luo, Z.-X., Zhang, X., Yuan, C.-X. & Xu, L. *Science* **326,** 278–281 (2009).
6. Simpson, G. G. *Am. J. Sci.* **15,** 461–470 (1928).

SIGNAL TRANSDUCTION

How cells sense energy

Maintaining an optimal cellular energy status requires sensing the levels of the adenine nucleotides ATP, ADP and AMP. Biochemical and structural studies of the enzyme AMPK provide insights into how this is achieved. SEE LETTER P.230

D. GRAHAME HARDIE

Most cellular processes consume energy, which is provided by the hydrolysis of adenosine triphosphate (ATP) to adenosine diphosphate (ADP). Cells usually maintain ATP:ADP ratios that are orders of magnitude away from equilibrium, in much the same way that chemicals in an electrical battery store energy. Animal cells recharge this 'battery' by oxidizing fuels such as glucose, using some of the released energy to convert ADP back to ATP. Clearly, it is essential that the battery remains constantly charged, and this requires systems that monitor the cellular levels of ATP and ADP as well as AMP (adenosine monophosphate, derived from ADP). Pre-eminent among these systems is AMP-activated protein kinase (AMPK) — a control enzyme that triggers corrective changes in metabolism when energy status is compromised[1]. On page 230 of this issue,

Xiao *et al.*[2] provide intriguing biochemical and structural insights into the mechanism by which cellular energy status is sensed. Their data suggest that AMPK responds not only to AMP and ATP as previously thought, but also to ADP.

AMPK consists of a catalytic α-subunit and two regulatory subunits (β and γ)[1]. On attachment of a phosphate group to the side chain of T172 (an amino-acid residue on the α-subunit), AMPK's catalytic activity increases more than 100-fold. But as upstream kinase enzymes catalyse phosphate attachment, under normal conditions phosphatase enzymes immediately remove it. AMP binding inhibits phosphate removal[3], providing a sensitive mechanism for switching on AMPK. AMP binding also promotes a further 10-fold increase in AMPK activity (yielding an overall activation of greater than 1,000-fold) by allosteric means — that is, by causing a change in the shape of AMPK that indirectly affects its catalytic activity[4].

The AMPK γ-subunit contains four tandemly arranged sequence repeats that form the binding sites for AMP, ADP or ATP[5]. An earlier structure[6] revealed that these repeats assemble symmetrically into a flattened disk, with one repeat in each quadrant; the disk contains four cavities in which the adenine nucleotides might bind. Of these, site 2 always seems to be unoccupied, whereas site 4 contains a permanently bound AMP of unknown function. As for sites 1 and 3, AMP binds to them reversibly, in competition with ATP and ADP.

Xiao *et al.*[2] first investigated adenine nucleotide binding to AMPK. They found that AMP, ADP and ATP bind with similar affinity to site 1, which seems to mediate the allosteric effect, and that this site has a 50-fold higher affinity for all three nucleotides than site 3, which regulates T172 dephosphorylation. The authors also made the intriguing finding that ADP protects against phosphate removal just as well as does AMP, although only AMP causes allosteric activation as reported previously[7].

In a tour de force of crystallography, Xiao *et al.* also provide the most complete AMPK structure to date: it contains almost the entire α-subunit, the carboxy-terminal domain of the β-subunit and the entire γ-subunit with AMP in sites 3 and 4. The structure shows that the
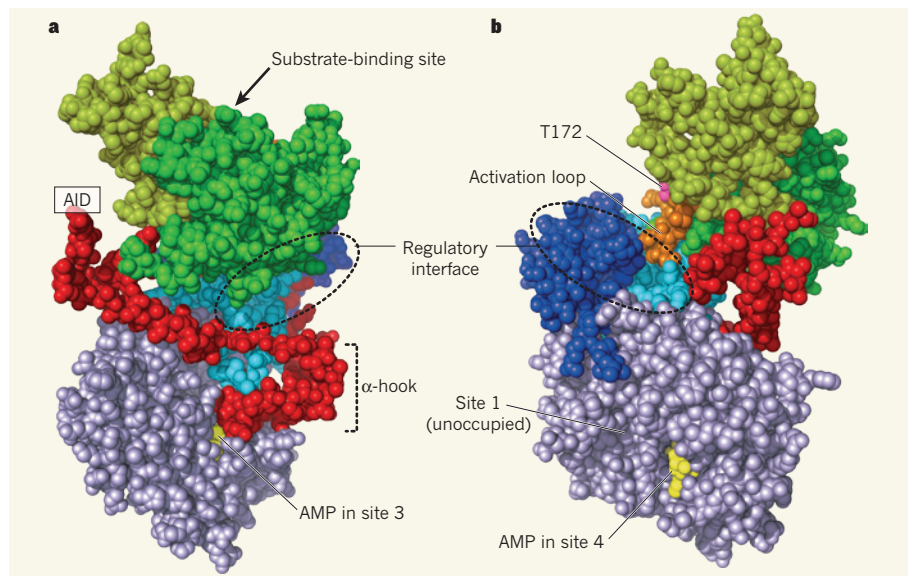
Figure 1 | The new AMPK structure from two angles[2]. a, The two lobes of the catalytic domain of the enzyme's α-subunit are shown in two shades of green. The arrow indicates the substrate-binding site at the back of the complex. The carboxy termini of the α-subunit (dark blue) and the β-subunit (light blue) together form the regulatory interface. The linker peptide of the α-subunit (red) contains a structure called the α-hook, which interacts with AMP (yellow) at the γ-subunit (purple), specifically at site 3. Although the structure of the auto-inhibitory domain (AID) in the α-subunit was not resolved in the crystallographic structure, its expected position is labelled. **b,** View from the left side of **a**. In the activation loop (orange) within the α-subunit, T172 (pink) lies in a cleft between the catalytic domain and the regulatory interface, where phosphatases would have restricted access.

catalytic (kinase) domain abuts a 'regulatory interface' (formed from the carboxy-terminal domains of the α- and β-subunits), with its substrate-binding site facing away from the rest of the complex (Fig. 1a). The activation loop, which houses T172, is well resolved in the structure, and interacts with the regulatory interface; this would restrict the access of phosphatases to phosphorylated T172 (Fig. 1b).

Xiao and colleagues' structure also provides an elegant explanation for how AMP binding to site 3 prevents T172 dephosphorylation. In the α-subunit, a long linker peptide that connects the catalytic and carboxy-terminal domains includes a structure termed the α-hook; residues at the tip of the hook interact with AMP in site 3 (Fig. 1a). The authors propose that a favourable interaction between the α-hook and AMP (or ADP) in site 3 stabilizes the binding of the catalytic domain to the regulatory interface, thus blocking phosphatase access to T172. They also suggest that, when ATP is bound, this interaction would be disrupted, causing the catalytic domain to dissociate from the regulatory interface and exposing T172 for dephosphorylation. Their biochemical analysis supports this model.

Naturally, puzzles remain. For instance, the structure provides no clues to how AMP binding to site 1 might allosterically activate the catalytic domain. It has been proposed[8] that this involves the auto-inhibitory domain (AID, residues 305–330), which follows the catalytic domain on the α-subunit. In the new structure no residues of the AID were resolved, suggesting that it was unfolded and mobile. Although it is possible that the AID folds up when AMP binds to site 1, they are quite remote from each other (Fig. 1b), making it difficult to see how the two events would be connected. What's more, a previously proposed idea[9] that AMP promotes phosphate attachment to — as well as inhibiting phosphate removal from — T172 was recently resurrected[10]. The new structure gives no indication of how this might happen.

Perhaps the most interesting finding of the paper[2] is that ADP prevents phosphate removal, which seems likely to be of physiological relevance. Normally, cellular concentrations of AMP are much lower than those of ADP, except in situations of severe energy deficit. During mild energy stress, displacement of ATP by ADP at site 3 would activate AMPK by preventing phosphate removal from T172; the allosteric effect of AMP binding at site 1, however, would amplify this response during more severe stress. This dual mechanism would allow AMPK to sense energy deficit progressively over a wide range of energy availability. ∎

D. Grahame Hardie *is in the College of Life Sciences, University of Dundee, Dundee DD1 5EH, UK.*
e-mail: d.g.hardie@dundee.ac.uk

1. Hardie, D. G. *Nature Rev. Mol. Cell Biol.* **8**, 774–785 (2007).
2. Xiao, B. *et al. Nature* **472**, 230–233 (2011).
3. Davies, S. P., Helps, N. R., Cohen, P. T. W. & Hardie, D. G. *FEBS Lett.* **377**, 421–425 (1995).
4. Suter, M. *et al. J. Biol. Chem.* **281**, 32207–32216 (2006).
5. Scott, J. W. *et al. J. Clin. Invest.* **113**, 274–284 (2004).
6. Xiao, B. *et al. Nature* **449**, 496–500 (2007).
7. Carling, D., Clarke, P. R., Zammit, V. A. & Hardie, D. G. *Eur. J. Biochem.* **186**, 129–136 (1989).
8. Chen, L. *et al. Nature* **459**, 1146–1149 (2009).
9. Hawley, S. A. *et al. J. Biol. Chem.* **270**, 27186–27191 (1995).
10. Oakhill, J. S. *et al. Proc. Natl Acad. Sci. USA* **107**, 19237–19241 (2010).

BIOMECHANICS

Swimming in the Sahara

The sandfish, a type of desert lizard, can vanish into a sandy substrate in a blink of an eye. Approaches that draw on mathematics, physics and engineering provide complementary insights into how the animal achieves this feat.

STEPHANIE B. CROFTS & ADAM P. SUMMERS

Biomechanists, with help from Isaac Newton and William Froude, have developed a good understanding of the forces involved in locomotion on solid ground[1]. Even if an animal should slip beneath the waves and undulate into the deep there are equations and techniques, courtesy of Claude-Louis Navier, George Stokes and computational fluid dynamics, that nicely explain and predict its motion[2].

However, when an animal moves through a granular material, predictability and theoretical underpinning have been as elusive as seeing a world in a grain of sand. In a paper in the *Journal of the Royal Society Interface*, Maladen and his colleagues[3] describe how they have used a refinement of their resistive force theory to show that a sand-swimming lizard moves forward about as fast as it can. Their model will allow biologists and engineers to explore locomotion in granular solids with unprecedented ease and speed.

Sandfish (*Scincus scincus*), also known as skinks, are not fish but lizards. They are drab, tan-coloured creatures with short, sprawling limbs. But they have one impressive ability —when startled, a sandfish can vanish completely into a North African or Middle Eastern dune in less than a second (Fig. 1). If you dig at the spot where it vanished you will find nothing but sand, the lizard having wriggled away under the surface.

In earlier work using X-ray videography,

Maladen and colleagues[4] found that sandfish move not by paddling with their legs but by undulating their bodies from side-to-side to swim through the sand. The sandfish throws its body into an S-shape with an amplitude of about 20% of its length. This sinusoidal wave travels backwards along the body at about twice the speed with which the animal moves forward, amounting to a wave efficiency of about 0.5. In their new work, the researchers used this conclusion as a benchmark for further studies with simulations, mathematical models and a robotic lizard.

The behaviour of a granular material such as sand is tricky to understand because it can act as both a solid and a fluid. The many small particles can flow like water but, under the right conditions, sand can be as unyielding as the rock from which it came. This is where resistive force theory (RFT) comes in.

The theory was inspired by models of tiny aquatic organisms moving in water, where movement is constrained by the viscosity of the fluid. In granular materials, however, it turns out that frictional drag is more important than viscosity, and drag values for short segments of the submerged object are an important input of RFT. One innovation of this work[3] is the use of drag values derived from simulations rather than those laboriously gathered by pulling objects through media of differing density. Using the simulation-derived drag values and kinematics measured from videos, Maladen *et al.*[3] found that RFT predicted the wave efficiency reasonably well, although not perfectly.

Figure 1 | Going, going, almost gone. *Scincus scincus* dives head first into the sand.

An alternative to RFT is a numerical modelling technique that takes into account the interactions of spherical grains with each other and with a deforming, submerged solid shaped like a sandfish. However, simulating sand-sized particles would take weeks of computer time. Fortunately, sandfish are happy to burrow in glass beads 3 millimetres in diameter. By observing the lizards moving through the beads, and then using the larger particle size in the simulations, Maladen *et al.* cut computing time to 'mere' days. The agreement between the simulation and the lizards' performance was better than that derived by RFT, but the procedure was much more time-consuming.

The advantage of both the RFT and numerical modelling techniques is that parameters can be varied to explore patterns and seek optima. Initially, the most striking result was the finding that wave efficiency continues to rise as the animal's body is thrown into higheramplitude waves. But these sharp bends mean that the sandfish covers less forward distance with each undulation. When the forward speed was plotted against the ratio of amplitude to wavelength, the peak speeds occurred at ratios ranging from 0.19 to 0.27 for the various simulations. The authors found that the data collected from actual lizard performance lie at this same peak. Thus, it seems that when fleeing from a potential predator, kinematic efficiency takes a back seat to a speedy retreat.

Designing a biomimetic robot teaches you as much about biology as it does about engineering[5,6]. In this case, Maladen and colleagues[3] used a sand-swimming robot to determine whether the theory, and the practice as demonstrated by the lizard itself, could be translated into the realm of engineering. The robot consisted of six connected segments, powered by servos, packed in a latex sock and wrapped in a spandex swimsuit.

The researchers showed that the robot is able to swim in granular media at similar speeds to those observed in the other models and in live organisms, but at efficiencies 30% below the predicted maximum. By simulating the segmented robot in the numerical model, they determined that this discrepancy in performance was due to the robot's low number of intersegmental joints. It was not until the researchers used more than 15 segments, creating a robot capable of forming a smoother curve, that wave efficiency approached a maximum. Perhaps the reason that many sand swimmers are elongate, smooth-bodied and small-limbed, with many vertebrae, is to better fall into a curve without the sharp bends that sap energy. ∎

Stephanie B. Crofts and **Adam P. Summers** *are at the University of Washington, Friday Harbor Laboratories, Friday Harbor, Washington 98250, USA.* e-mails: croftss@uw.edu; fishguy@uw.edu

1. Dickinson, M. H. *et al. Science* **288,** 100–106 (2000).
2. Lauder, G. V. & Tytell, E. D. in *Fish Biomechanics* (eds Shadwick, R. E. & Lauder, G. V.) 425–468 (Academic, 2006).
3. Maladen, R. D., Ding, Y., Umbanhowar, P. B., Kamor, A. & Goldman, D. I. *J. R. Soc. Interface* doi:10.1098/rsif.2010.0678 (2011).
4. Maladen, R. D., Ding, Y., Li, C. & Goldman, D. I. *Science* **325,** 314–318 (2009).
5. Lentink, D. *et al. Nature* **446,** 1082–1085 (2007).
6. Birdwell, J. A. *et al. J. Neurophysiol.* **98,** 2439–2455 (2007).

SPECTROSCOPY

A closer look at polymer annealing

Solvent vapour annealing processes are used to optimize the material properties of thin films of semiconducting polymers used in electronic devices. One such process has now been examined at the molecular level.

YI FU & JOSEPH R. LAKOWICZ

Single-molecule spectroscopy and microscopy (SMS) techniques have had a tremendous impact in biophysics — in studies of the motion of individual protein molecules, for example. SMS techniques are also suitable for investigating the underlying heterogeneity in complex liquids and solids, such as that associated with different molecular conformations, and for directly observing dynamic changes in materials at the molecular level. Polymeric materials would therefore seem a natural fit for single-molecule studies, as they are, by definition, structurally heterogeneous. Yet SMS has so far not been widely used in polymer science, apart from in studies of certain fluorescent polymers[1–4].

Reporting in *Angewandte Chemie*, Vogelsang *et al.*[5] now describe the use of SMS techniques to reveal the conformational dynamics of single polymer chains during solvent vapour annealing (SVA) — an industrially important process in which the exposure of polymer films to solvent vapour enables the polymer molecules to reach conformational equilibrium, thus optimizing the films' useful properties. An understanding of SVA at the molecular level should lead to improvements in this widely used process.
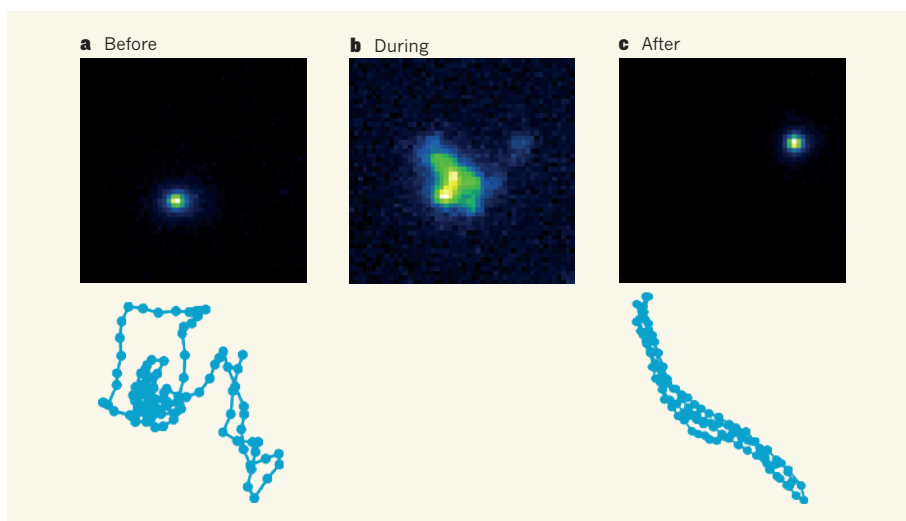
Figure 1 | Monitoring single polymeric chains. Vogelsang *et al.*[5] have used single-molecule spectroscopy and microscopy to study the conformational changes of individual MEH-PPV polymer molecules isolated in a thin film of PMMA (another polymer) before, during and after solvent vapour annealing. Each of the panels **a** to **c** shows accumulated images, taken over 60 seconds, of the same MEH-PPV chain embedded within an 8×8-micrometre region of PMMA film. **a**, Before annealing, single MEH-PPV chains are visible as static fluorescent spots, which correspond to collapsed, disordered polymeric conformations (such as the structure shown beneath the fluorescence data). **b**, During annealing, the spots are mobile and the background fluorescence increases. Here, the fluorescent spot is extended and blurred because of the motion of the single chain as it moves from left to right. **c**, After annealing, static spots are again visible, but these now correspond to extended, highly ordered polymer conformations, such as the one shown. (Figure adapted from ref. 5.)

Annealing processes are attractive techniques for improving the material properties of organic semiconductors (such as conductive polymers) for use in transistors, solar cells and other devices[6]. SVA induces translocations and folding/unfolding of molecules in polymer films at ambient temperatures, eventually leading to the formation of macrostructures featuring equilibrated, long-range molecular order[7,8]. But so far, our understanding of the morphological transformations — the changes in polymer-chain conformations — that occur during SVA has been sketchy.

Vogelsang *et al.*[5] are the first to use SMS to study real-time conformational changes in single conductive polymer chains during SVA. They investigated a prototype conductive polymer (poly[2-methoxy-5-(2′-ethylhexyloxy)-1,4-phenylene vinylene]; MEH-PPV) isolated as single molecules in a host matrix (a non-conductive polymer, poly(methyl methacrylate); PMMA). The authors observed clear differences in the translational diffusion of single MEH-PPV chains before, during and after SVA. Before and after annealing, single chains were visible as sharp, stationary fluorescent spots, but during annealing, the spots were blurred and danced around freely (Fig. 1).

From these results, Vogelsang *et al.* concluded that MEH-PPV/PMMA films absorb solvent and swell during SVA, and that the annealing process lowers the glass-transition temperature of the film to below ambient temperature, which increases chain mobility. The film therefore exists as a heterogeneous mixture of solid and liquid-like phases during annealing. Crucially, the authors were able to monitor the translational motion of single MEH-PPV chains in the solid and liquid-like phases by simply tracking the movement of the fluorescence spots. Some of the single-chain diffusion was, however, too rapid to be observed using microscopy, but the authors were able to detect this using a technique known as fluorescence correlation spectroscopy. The authors' results also showed that SVA generates different kinds of swelling processes in the host matrix, and causes single MEH-PPV polymer chains to diffuse in several different ways.

In the past few years, various studies[1–3] have provided evidence that MEH-PPV molecules adopt numerous configurations within a PMMA host, ranging from complex disordered structures that are roughly cylindrical, to highly ordered ones such as toroids and rods. The three-dimensional shape of MEH-PPV polymer chains is expected to have a profound impact on the electronic properties of the molecules and on their individual fluorescence quantum yields — a measure of how much light the chains produce in response to excitation by incoming photons. Vogelsang and colleagues[5] found that MEH-PPV polymer chains undergo folding and unfolding events between collapsed and extended conformations during SVA. The transition between these conformations correlates with fluctuations in the fluorescence intensity of single chains, which occur because the fluorescence quantum yield of MEH-PPV is lower in collapsed conformations.

The authors also characterized the conformational order of single MEH-PPV chains using fluorescence excitation polarization spectroscopy, which measures variations in fluorescence in different directions. From these data, Vogelsang *et al.* concluded that SVA equilibrates MEH-PPV single chains in the matrix towards lower-energy conformations, which ultimately leads to the production of highly ordered conformations after annealing. The authors used the same technique to analyse vapour-annealed MEH-PPV/PMMA films that had been prepared in different solvents. They observed that, as long as SVA was performed using a good solvent (one in which the film is soluble), then single chains could be reorganized into low-energy, high-order conformations regardless of the solvents in which the film was prepared.

The ability to follow annealing effects at the molecular level will undoubtedly improve our understanding of polymer reorganization processes, and offers a complementary approach to microscopy methods that monitor the surface features and textures of thin films. Insights into how single-chain morphology changes in thin films, block co-polymers, or even in highly ordered self-assembled structures, will help to direct molecular-design strategies for making improved conductive polymers for device applications. Vogelsang and colleagues' findings are obviously of relevance to industrially important processes for assembling conductive polymer thin films, but in the longer term, they might also be applicable to the control of morphological order in other polymeric functional materials. We expect that surprising discoveries will be made as single-molecule methods are used to explore other subjects in polymer science. ∎

Yi Fu *and* **Joseph R. Lakowicz** *are in the Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA.* e-mails: yifu@cfs.umaryland.edu; lakowicz@cfs.umaryland.edu

1. Adachi, T. *et al. J. Phys. Chem. C* **114,** 20896–20902 (2010).
2. Barbara, P. F. *et al. Acc. Chem. Rev.* **38,** 602–610 (2005).
3. Gesquiere, A. J., Lee, Y. J., Yu, J. & Barbara, P. F. *J. Phys. Chem. B* **109,** 12366–12371 (2005).
4. Muls, B. *et al. ChemPhysChem* **6,** 2286–2294 (2005).
5. Vogelsang, J., Brazard, J., Adachi, T., Bolinger, J. C. & Barbara, P. F. *Angew. Chem. Int. Edn* **50,** 2257–2261 (2011).
6. Albert, J. N. L. & Epps, T. H. *Mater. Today* **13,** 24–33 (2010).
7. Conboy, J. C. *et al. J. Phys. Chem. B* **102,** 4516–4525 (1998).
8. Dickey, K. C., Anthony, J. E. & Loo, Y. L. *Adv. Mater.* **18,** 1721–1726 (2006).

HOST–MICROBE INTERACTION

Innate immunity cues virulence

Salmonella **intestinal pathogens employ a clever trick. They use the immune response that their host triggers to destroy them to enhance the expression of genes that mediate the pathogens' virulence.**

MICHELLE M. C. BUCKNER & B. BRETT FINLAY

Bacterial pathogens use complex strategies to survive and replicate, causing disease as they do so. But how these strategies, which are usually mediated by molecules called virulence factors, are regulated during infection is poorly understood. In a paper published in *Cell*, Arpaia *et al.*[1] elegantly demonstrate that the Typhimurium serovar of the bacterium *Salmonella enterica* activates the host immune system and then uses this innate response as a signal to induce its own virulence genes.

Salmonella species are Gram-negative intestinal pathogens that cause disease ranging from gastroenteritis to typhoid fever, which can be fatal[2]. They are transmitted mainly through contaminated food or water, and colonize their host's intestine. In typhoid fever, the pathogens can cross the intestinal barrier and spread to the spleen and liver by entering and replicating in phagocytic immune cells such as macrophages[2].

The innate immune system is crucial for controlling infectious agents. It recognizes general pathogen-associated molecular patterns (PAMPs), and its function is to both kill pathogens and alert the adaptive immune system[3]. Integral to innate immunity is a family of proteins called Toll-like receptors (TLRs), which recognize various PAMPs and initiate signalling cascades that act to protect the host[3].

Mammals have many TLRs, each of which recognizes certain PAMPs. For example, TLR4, found on the cell surface, recognizes
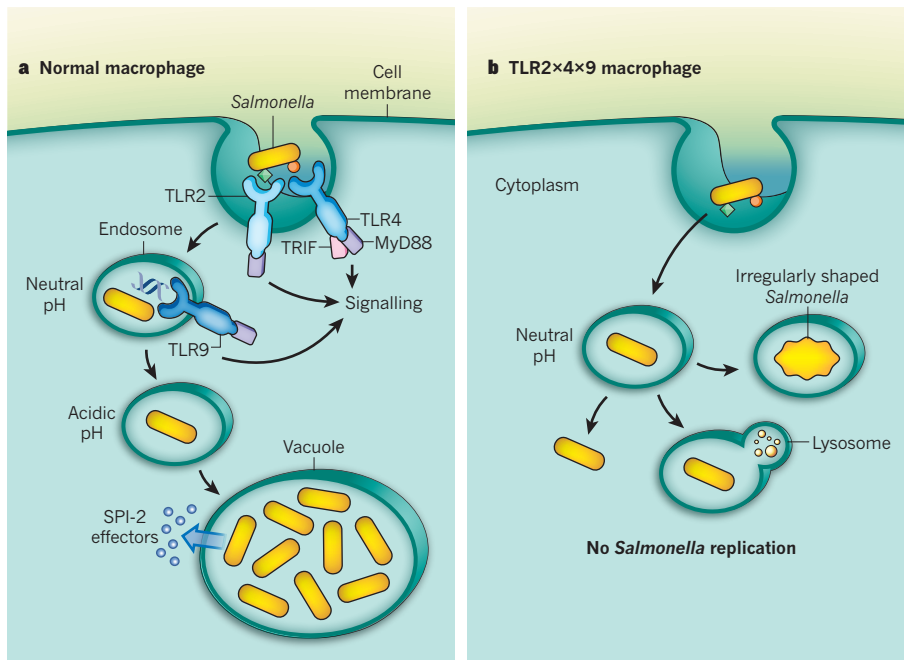
Figure 1 | *Salmonella* and its effect on macrophages. a, In normal macrophages, Toll-like receptors (TLRs) recognize pathogens such as *Salmonella* and, together with adaptor molecules including MyD88 and TRIF, initiate downstream signalling pathways to trigger appropriate immune responses. Arpaia *et al.*[1] show that activation of TLRs 2, 4 and 9 acidifies the vacuole containing engulfed *Salmonella*. This in turn induces the expression of bacterial SPI-2 genes and translocation of the effector molecules into the host-cell cytoplasm. SPI-2 activity transforms the vacuole into a replicative niche, allowing bacterial numbers to increase. **b**, In macrophages lacking these three TLRs (TLR2×4×9), the pH of the *Salmonella*-containing vacuole does not become acidic as rapidly or to the same extent, and SPI-2 is not induced. This has several consequences: bacterial release into the macrophage cytoplasm; vacuole fusion with lysosomes and so bacterial death; and irregularly shaped *Salmonella*, which correlates with decreased bacterial survival.

lipopolysaccharide, a component of the bacterial cell wall[3]. TLR2, which is also located in the cell membrane, recognizes bacterial surface lipoproteins. And TLR9, located in the membrane of cellular organelles called endosomes, recognizes non-methylated CpG sequences in bacterial DNA.

When a TLR binds its specific PAMP, host-cell adaptor molecules — such as MyD88 and TRIF — are recruited, and downstream signalling is initiated[3]. TLR signalling helps to resolve infection not just by activating specific bactericidal mechanisms, but also indirectly by inducing the production of pro-inflammatory proteins such as cytokines. The function of these receptors has generally been investigated in cell lines or in mice lacking MyD88 or TRIF. But this can be problematic, because MyD88 and TRIF also act in non-TLR-associated cellular pathways[1,3].

Arpaia *et al.*[1] show that *Salmonella* activates its virulence genes by using as a signal TLR-induced acidification of the phagosome — the intracellular compartment that encloses the engulfed pathogen. The molecules encoded by these genes allow the bacterium to replicate in the normally bactericidal phagosome.

The authors find that mice lacking TLRs 2, 4 and 9 (TLR2×4×9 mice) were less susceptible to *Salmonella* infection than those lacking only TLRs 2 and 4 or TLRs 4 and 9. Moreover,

bone-marrow-derived macrophages (BMMs) from TLR2×4×9 mice or MyD88×TRIF mice were more resistant to *Salmonella* infection than the equivalent cells from TLR2×4 mice. In fact, Arpaia *et al.* report that TLR2×4×9 BMMs can support the replication of several intracellular pathogens, but not *Salmonella*. These observations indicate the importance of TLR signalling in *Salmonella* virulence.

How does *Salmonella* exploit TLR signalling? During an infection, these pathogens normally transform the phagosome into a special vacuole, termed the *Salmonella*-containing vacuole (SCV), in which they replicate[2]. In TLR2×4×9 BMMs, however, SCV formation is impaired, and the bacteria are detected in the cytoplasm and associate with lysosomes — intracellular organelles that have bactericidal activity (Fig. 1).

Salmonella forms, and survives within, the SCV by activating a set of genes that occur in a locus within their genome termed SPI-2. These genes encode a type-III secretion system that functions as a syringe to translocate SPI-2 effector proteins from the bacterial cytoplasm into the host-cell cytoplasm[2,4]. The effectors manipulate the host cell to allow bacterial replication[2]. Arpaia *et al.* studied levels of SPI-2 gene expression and effector translocation and found that — unlike the case in normal or TLR2×4 BMMs — SPI-2 was not induced

in TLR2×4×9 cells. The authors could restore bacterial replication in TLR2×4×9 BMMs using a *Salmonella* mutant that constitutively expresses SPI-2 genes[5].

This paper[1] clearly shows that, rather than protecting against *Salmonella* infection, functional TLR signalling contributes to the pathogen's full virulence. Arpaia *et al.*[1] shed light on this apparent paradox by demonstrating that the vacuole of TLR2×4×9 and MyD88×TRIF BMMs does not acidify as rapidly or to the same extent as the SCV of normal and TLR2×4 BMMs. It is widely accepted[6] that *Salmonella* uses pH as a signal for SPI-2-gene induction and effector translocation. So it seems that the bacteria take advantage of the essential, and additive, effects of TLRs on vacuole acidification to modulate the expression of virulence genes.

The study also raises several questions. What are the exact pathway(s) linking TLRs and vacuolar acidification? Some SPI-2 effectors also manipulate aspects of the host immune system — for instance, the activity of the NF-κB transcription factor[7]. Such specific effectors might also modulate TLR signalling to induce the expression of virulence genes. Alternatively, more-complex regulatory steps may be involved.

And how do TLRs interact with intracellular *Salmonella*? Does *Salmonella* release specific ligands such as nucleic acids into the phagosome to activate TLRs? There is a precedent for this in other species[8]. So it is possible that *Salmonella* also uses such ligands to induce a robust immune response.

More broadly, do *Salmonella* and other bacteria exploit other host immune responses? Aparia and colleagues' work underlines the importance of collaboration between workers in immunology and microbiology. Together, these research areas are set to shed light on the intricate interactions between pathogens and their hosts. ∎

Michelle M. C. Buckner *and* **B. Brett Finlay** *are in the Department of Microbiology and Immunology, and at the Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada.*
e-mail: bfinlay@msl.ubc.ca

1. Arpaia, N. *et al. Cell* **144,** 675–688 (2011).
2. Haraga, A., Ohlson, M. B. & Miller, S. I. *Nature Rev. Microbiol.* **6,** 53–66 (2008).
3. Yamamoto, M. & Takeda, K. *Gastroenterol. Res. Pract.* **2010,** 240365 (2010).
4. Shea, J. E., Hensel, M., Gleeson, C. & Holden, D. W. *Proc. Natl Acad. Sci. USA* **93,** 2593–2597 (1996).
5. Silphaduang, U., Mascarenhas, M., Karmali, M. & Coombes, B. K. *J. Bacteriol.* **189,** 3669–3673 (2007).
6. Yu, X.-J., McGourty, K., Liu, M., Unsworth, K. E. & Holden, D. W. *Science* **328,** 1040–1043 (2010).
7. Le Negrate, G. *et al. J. Immunol.* **180,** 5045–5056 (2008).
8. Woodward, J. J., Iavarone, A. T. & Portnoy, D. A. *Science* **328,** 1703–1705 (2010).

ARTICLE

Transitional mammalian middle ear from a new Cretaceous Jehol eutriconodont

Jin Meng[1,2], Yuanqing Wang[2] & Chuankui Li[2]

The transference of post-dentary jaw elements to the cranium of mammals as auditory ossicles is one of the central topics in evolutionary biology of vertebrates. Homologies of these bones among jawed vertebrates have long been demonstrated by developmental studies; but fossils illuminating this critical transference are sparse and often ambiguous. Here we report the first unambiguous ectotympanic (angular), malleus (articular and prearticular) and incus (quadrate) of an Early Cretaceous eutriconodont mammal from the Jehol Biota, Liaoning, China. The ectotympanic and malleus have lost their direct contact with the dentary bone but still connect the ossified Meckel's cartilage (OMC); we hypothesize that the OMC serves as a stabilizing mechanism bridging the dentary and the detached ossicles during mammalian evolution. This transitional mammalian middle ear narrows the morphological gap between the mandibular middle ear in basal mammaliaforms and the definitive mammalian middle ear (DMME) of extant mammals; it reveals complex changes contributing to the detachment of ear ossicles during mammalian evolution.

The lower jaw of non-mammalian amniotes is composed of the tooth-bearing dentary and several post-dentary bones; that of mammals is formed by the dentary alone. In contrast, there is only one ossicle, the columella auris (or stapes), in the middle ear of non-mammalian amniotes, but there are multiple ossicles in mammals, including the malleus, incus, stapes and ectotympanic[1–4]. Fossils have shown a series of reductions of the post-dentary bones during synapsid evolution towards mammals[1,2,4–6], and developmental studies have demonstrated homologies of mammalian middle ear ossicles with their reptilian precursors, including the malleus (=articular plus prearticular), incus (=quadrate) and ectotympanic (=angular)[7–9]. In basal mammaliaforms, such as *Morganucodon*, the post-dentary bones have greatly reduced but still attach to the dentary, serving a dual function for hearing and feeding[2,4,5,10,11]. The mandibular middle ear of *Morganucodon*[12] is regarded as the prototype that gives rise to the definitive mammalian middle ear (DMME) in which the angular, articular plus prearticular, and quadrate are strictly auditory structures and fully divorced from the feeding apparatus[2,4]. Incorporation of the lower jaw elements and the quadrate into the middle ear on the cranium represents an innovative feature of mammals and has been regarded as a classic example of gradual evolution in vertebrates, a subject that has attracted enormous attention[1–19]. Conventional research on the evolution of the mammalian middle ear has focused primarily on detachment of the post-dentary bones[2–4,10,11,14,16–18,20], in which transitional changes are often inferred from grooves on the medial surface of the dentary[4,6] or from fragmentary specimens[18]. The key questions still remain: what has happened, and how did it happen, during the transference from the mandibular middle ear to the DMME?

In view of the middle ear from the new Cretaceous eutriconodont mammal reported here, we hypothesize that the ossified Meckel's cartilage (OMC) serves as a stabilizing mechanism for the post-dentary bones during their evolutionary departure from the dentary. We define the transitional mammalian middle ear (TMME) as a distinct stage in the evolution of the mammalian middle ear and characterize the transference with a suite of morphological changes.

Mammalia Linnaeus, 1758
Eutriconodonta Kermack *et al.*, 1973
Liaoconodon hui, gen. et sp. nov.

Etymology. *Liao*, the short form for Liaoning Province; *conodon*, Latin, 'cuspate tooth', a common suffix for generic names in 'triconodonts'; *hui*, after Yaoming Hu, a student dedicated to the study of Mesozoic mammals.

Holotype. IVPP (Institute of Vertebrate Paleontology and Paleoanthropology, Beijing) V16051, a skeleton preserved on one slab of laminated siltstone (Figs 1 and 2 and Supplementary Figs 1–6).

Locality and horizon. Xiao-tai-zi, La-ma-dong, Jianchang, Liaoning, China; Jiufotang Formation. Early Cretaceous, Aptian, 120 Myr (ref. 21) (Supplementary Information part A).

Diagnosis. A medial-sized eutriconodont with a body length of 195 mm (from the tip of the rostrum to the hip) or 357 mm including the tail (Fig. 1). Dental formula I3.C1.P2.M3/i2.c1.p2.m4, with molariform teeth having three main cusps mesiodistally arranged and laterally compressed, and a dentary lacking an angular process and post-dentary trough. Differs from all known eutriconodonts in having the lower incisors, canine and first premolariform proportionally enlarged, similar in shape, and closely packed in space, and in having a deep trench along the ventral border of the masseteric fossa. Further differs from triconodontids in having the main cusp A/a distinctively higher and more inflated than other cusps; from 'jeholodentids'[18,22] in being larger and having relatively smaller (shorter) molariforms and a different dental formula; from gobiconodontids[12,23,24] in having fewer molariforms and in lacking a proportionally enlarged first incisor and the interlocking mechanism between successive lower molariforms; from *Repenomamus*[25,26] in being considerably smaller, having a different dental formula and less inflated cusp A/a in molariforms; from 'amphilestids' in having fewer molariforms and cheek teeth asymmetrical in labial view. Phylogenetic analyses place *Liaoconodon* between triconodontids and paraphyletic 'jeholodentids' (Supplementary Information and Supplementary Fig. 8).

Description

The description focuses on the middle ear region (Fig, 2, Supplementary Information part B and Supplementary Figs 2–6). The mandibular condyle is robust but not rounded. The glenoid fossa is well defined.

[1]Division of Paleontology, American Museum of Natural History, Central Park West at 79th street, New York, New York 10024, USA. [2]Key Laboratory of Evolutionary Systematics of Vertebrates, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, P.O. Box 643, Beijing 100044, China.
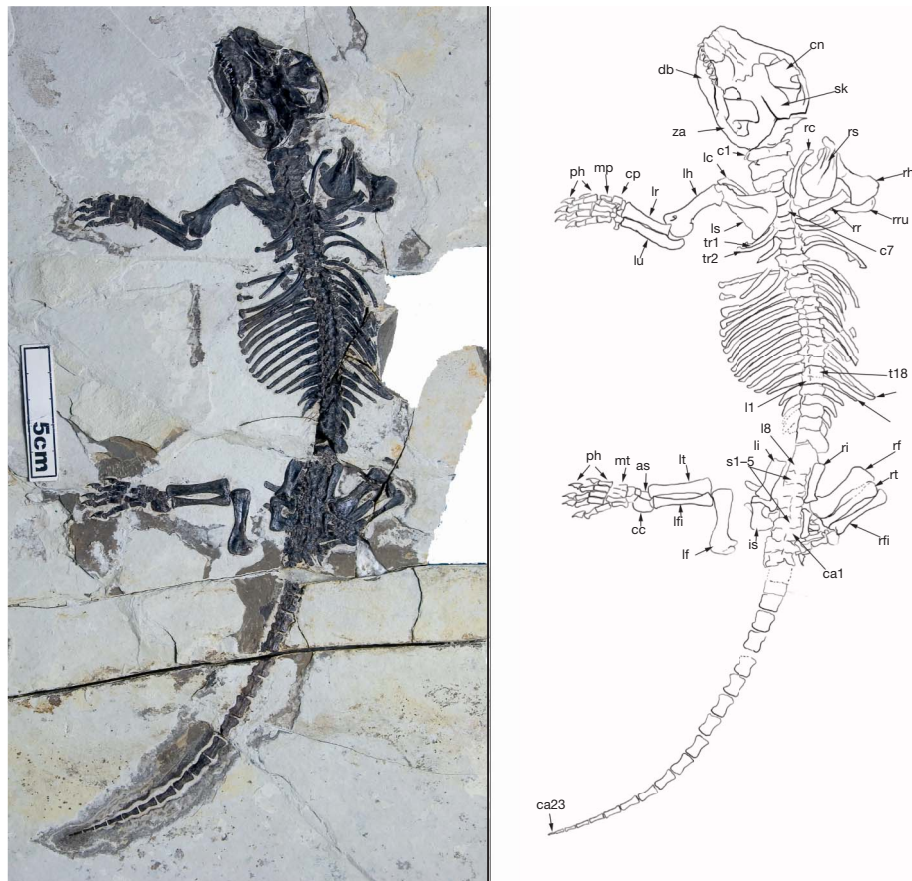
Figure 1 | The skeleton of the new mammal *Liaoconodon hui* (dorsal view of the holotype, IVPP V16051, Institute of Vertebrate Paleontology and Paleoanthropology, Beijing). Abbreviations: as, astragalus; c1–7, first to seventh cervical vertebrae; ca1–23, first to twenty-third caudal vertebrae; cc, calcaneum; cn, coronoid; cp, carpals; db, dentary bone; l1–8, first to eighth lumbar vertebrae (the posterior ones are damaged); lc, left clavicle; lh, left humerus; li, left ilium; lr, left radius; ls, left scapula; lu, left ulna; mp, metacarpals; mt, metatarsals; ph, phalanges; rc, right clavicle; rf, right femurs; rfi, right fibular; rh, right humerus; ri, right ilium; rr, right radius; rs, right scapula; rt, right tibia; ru, right ulna; s1–5, first to fifth sacral vertebrae; sk, skull; tr, thoracic ribs; za, zygomatic arch.

Immediately medial to the glenoid fossa is the epitympanic recess. The squamosal sulcus[24] is a short and narrow valley between the postglenoid process and the paroccipital process (Fig. 2). The medial surface of the dentary has the Meckelian groove that bifurcates posteriorly, similar to that of *Amphitherium*[4]. The OMC[16,17] is splint-like anteriorly and thickened posteriorly with the posterior portion bending medially as in other species[16–18,27]. The three-pronged ectotympanic (=angular) abuts the malleus laterally and is similar to that of extant mammals in embryonic stage[9,20,28–31]. Its anterior limb contacts the prearticular and the posterior limb ends lateral to the malleo-incudal articulation. The ventral limb of the ectotympanic (=the reflected lamina of the angular) is crescent shaped and is less developed (shorter) than that in extant mammals (Fig. 2), indicating that the ectotympanic can hold only the anterior portion of the tympanic membrane (Figs 2 and 3). As in multituberculates[32] and extant mammals[33–35], the inner surface of the curved bone bears the tympanic sulcus for attachment of the tympanic membrane.

The malleus consists of a long anterior process and a hook-shaped body, homologous with the prearticular and the articular, respectively[2,4,7–9]. In the medial view of the bone, a zigzag suture exists between the two elements (Fig. 2). A longitudinal groove is present in the anterior portion of the anterior process, perhaps for the passage of the chorda tympani nerve; there is no foramen for the latter. The body of the malleus is probably equivalent to the pars transversalis in extant mammals[14,33,34]; it has the neck and the manubrial base but lacks the manubrium. We consider the ventral portion of the malleus body homologous with the retroarticular process of mammaliaforms[2,10,11]. The ventromedial edge of the body is rugose, indicating attachment of soft tissue in life, probably for a sizeable tensor tympani homologous

with the posterior pterygoid muscle[36]. On the posterodorsal corner of the bone is a process that is absent on the malleus of extant mammals[33,35]. The morphology and bone relationships show this process to be most probably homologous with the surangular boss in advanced cynodonts[2,37]. The lateral (ventral) surface of the malleus is flat and smooth, suggesting contact with the tympanic membrane in life. The incus is posterior to the malleus and free from the skull; it has a convex articular facet for the malleus and a process presumably for articulation with the stapes. The malleo-incudal articulation is hinge-like. The lateral surface of the incus is smooth and even with the lateral surface of the malleus if the bones are in articulation. Bone fragments at the fenestra ovalis (vestibuli) may represent parts of the stapes, and, if true, the stapes must be relatively massive.

The OMC as an ossicle stabilizer

The OMC in *Yanoconodon*, another euriconodont from the Jehol Biota, was interpreted as having a paedomorphic resemblance to the embryonic pattern of modern mammals[18]. The middle ear of *Liaoconodon*, however, differs from that of *Yanoconodon* (Supplementary Information). The bone relationships (Fig. 2) show that during the early development of *Liaoconodon* the posterior end of the Meckel's cartilage must be separated from the rest of the cartilage, take the shape of the malleus body and join the dermal goniale (prearticular) to form the malleus; the rest of the cartilage became ossified at a later time but did not fuse with ear ossicles. In extant mammals the goniale usually ossifies earlier than other portions of the ossicles[30,31] and has a profound local influence upon endochondral ossification of the malleus[29]. Moreover, the Meckel's cartilage never fuses with the ectotympanic. These features
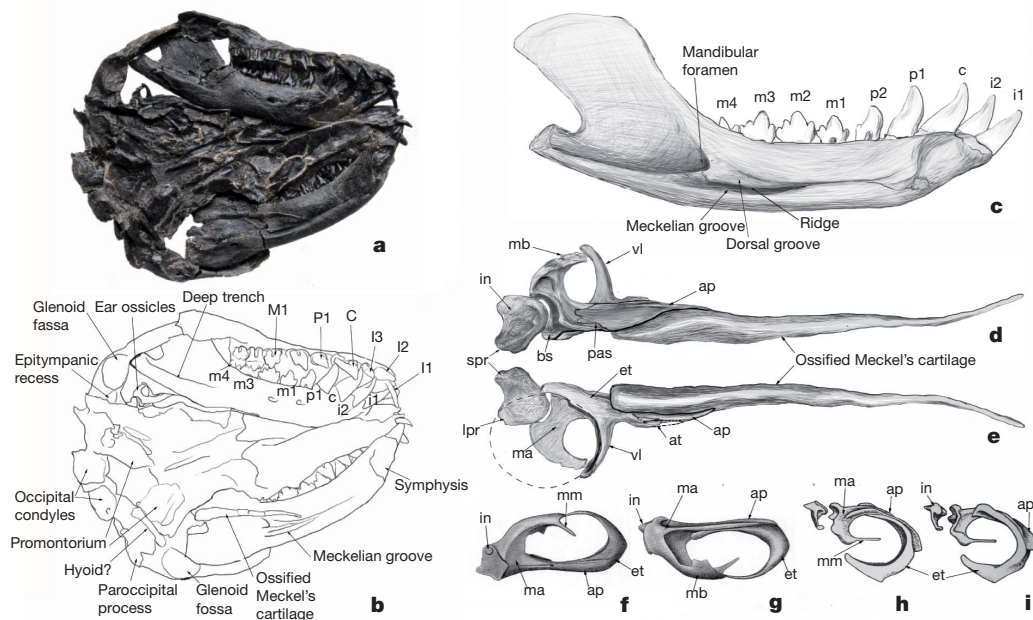
Figure 2 | Skull and ear ossicles of *Liaoconodon hui* (IVPP V16051).
a, **b**, Ventral view of the skull. **c**, Medial view of the dentary. **d**, **e**, Medial (dorsal) and lateral (ventral) views of the ossicles of *Liaoconodon*. Dashed oval indicates the estimated size of the tympanic membrane (actual shape could be more complicated). **f**, **g**, Dorsal and ventral views of the ossicles of *Ornithorhynchus anatinus* (modified after ref. 14). **h**, **i**, Medial and lateral views of the ossicles of *Didelphis* (modified after ref. 2). Ossicles are not on the same scale. The ear ossicles of *Liaoconodon* differ from those of extant mammals in having at least the following features: a long anterior process of the malleus wrapping around the OMC, ossicles proportionally larger, a hinge-like malleo-incudal joint, presence of the posterodorsal process of the malleus (boss of the surangular), lack of the manubrium, the ectotympanic partly developed. Abbreviations: ap, anterior process of malleus (prearticular); at, anterior process of the tympanic; bs, boss of surangular; et, ectotympanic (angular); in, incus (quadrate); lpr, long process of the incus; ma, body of malleus (articular); mb, manubrial base of malleus (retroarticular process); mm, manubrium of malleus; pas, prearticular–articular suture; spr, short process of the incus; vl, ventral limb of ectotympanic (reflected lamina of angular).

indicate that presence of the OMC in mammals[16–18,23,27] may not be interpreted as a simple paedomorphic feature.

It has been postulated that the Meckelian cartilage probably existed during a transitional stage in the evolution of extant mammals, later being replaced by the sphenomandibular ligament[1,4,5], and that such a ligament functions as a stabilizer for the detached ossicles[4]. Developmental studies also predict a persisting or ossified Meckel's cartilage in the common ancestor of recent mammals[14]. Along the same line of reasoning, we further hypothesize that during the evolution of mammals, the OMC or a persistent Meckel's cartilage functions as a stabilizing mechanism for the ear ossicles departed from the dentary but not yet supported by any cranial structure, as in the case of *Liaoconodon* (Fig. 3). Because the ossicles hold the tympanic membrane, it is necessary for them to be supported anteriorly so that the apparatus can function properly for hearing. This stabilizing mechanism becomes unnecessary when the ectotympanic fully suspends the tympanic membrane and the ossicles are moored on the cranium in more advanced mammals. In view of the new specimen, we think it inevitable to reinterpret the prearticular in the post-dentary unit of *Morganucodon*[11,12] as primarily the OMC (Fig. 3 and Supplementary Information part C).

Although known only in a few eutriconodonts[16–18,23] and symmetrodonts[17,27], the OMC or a persistent Meckel's cartilage in adults can be inferred from the Meckelian groove present in many early mammals (Fig. 4 and Supplementary Fig. 8). A Meckel's cartilage persisting in adult individuals of mammals has been observed or inferred elsewhere[14,38–40]. The sparse fossil record of the OMC is very probably because of its loose attachment to the dentary[16–18,23] so that preservation of the OMC is rare, similar to the record of ear ossicles[18,32,40]. It is also probable that the cartilage is not ossified; thus it left a contact groove on the dentary but is not fossilized.

Instead of being a paedomorphic resemblance, an alternative hypothesis is that the persistent Meckel's cartilage in Mesozoic mammals, along with features such as lack of the manubrium and a partial

ectotympanic, represents a phylogenetic stage in mammalian evolution, and that the embryonic pattern of modern mammals recapitulates the phylogenetic changes.
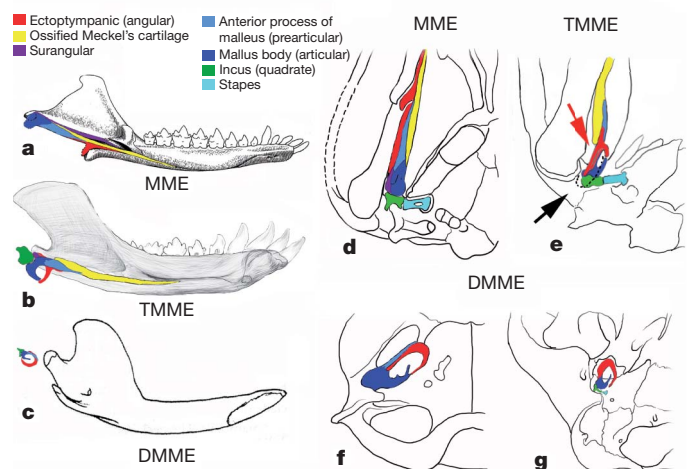
Figure 3 | Morphological transference of mammalian middle ear.
a–c, Medial views of the mandibles of *Morganucodon*, *Liaoconodon* and a generalized therian, showing the relationship with the OMC and ear ossicles. **d–g**, Ventral views of the basicranial and ear regions in *Morganucodon*, *Liaoconodon*, *Ornithorhynchus* and *Didelphis*, illustrating the relationship of the OMC, ear ossicles, the dentary and the basicranium. The black arrow in **e** points to the external auditory meatus, the red arrow to the gap between the ossicles and the dentary. The transference from the mandibular middle ear (MME) to the TMME and then to the DMME represents two distinct evolutionary stages, each involving several morphological changes (Supplementary Information part C). Source of drawings: **a** and **d** modified after refs 10 and 12, with the composition of bones reinterpreted; **c** from ref. 12; **f** from ref. 14; **g** from ref. 4.
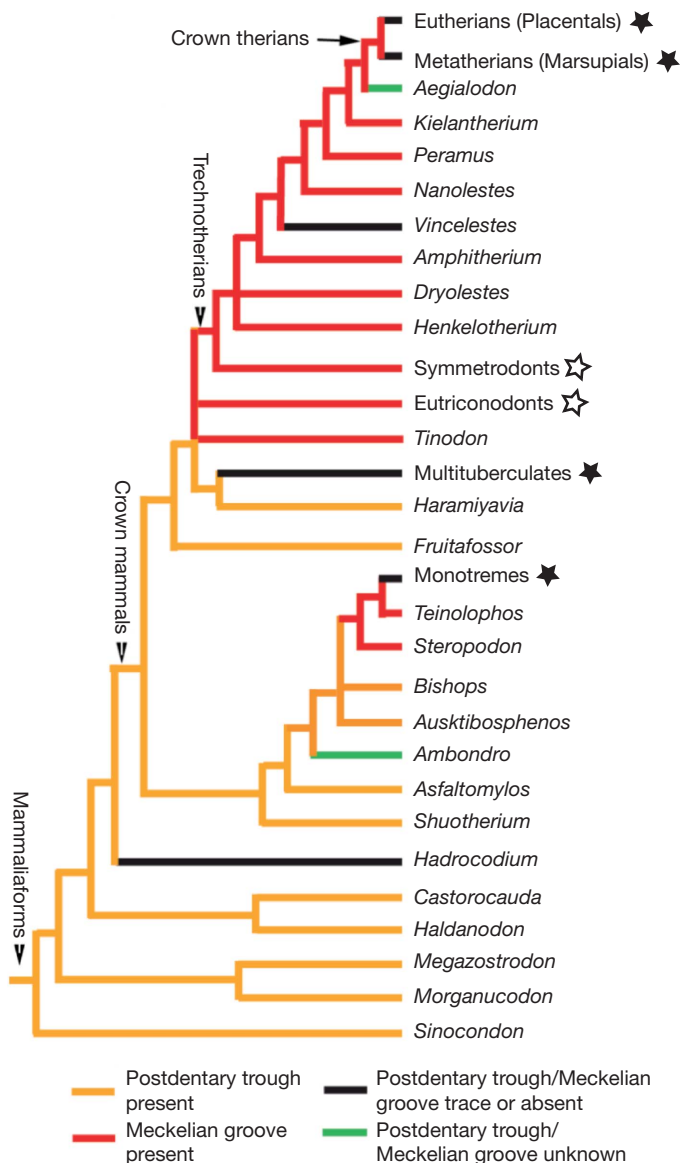
Figure 4 | Phylogeny of mammaliaforms (simplified after ref. 18) mapping distributions of the post-dentary trough and the Meckelian groove. A solid star denotes a group where the DMME is known in at least some members. An empty star indicates presence of the OMC is known in at least some members of the group. Given the phylogeny, the DMME evolved several times independently, and the transitional mammalian middle ear with a persistent Meckelian cartilage as a stabilizing mechanism, as inferred from presence of the groove, is consistent with the phylogeny (see also Supplementary Fig. 7).

The mammalian tympanic membrane

The origin of the mammalian tympanic membrane has been under extensive discussion[1,2,4,5,11,19,41–46] but still remains to be explained[5]. Four hypotheses have been outlined for the possible location and composition of the membrane in pre-mammalian synapsids, from which the mammalian condition is derived: (1) in a postquadrate location[1]; (2) in a post-dentary location[2]; (3) in both positions separately[44]; or (4) 1 and 2 coexisting as a continuous tympanic membrane[4]. The evidence from *Liaoconodon* supports hypothesis 4 (Fig. 4).

In *Liaoconodon* the crescent ectotympanic is insufficient to suspend the tympanic membrane fully; the latter must be in contact with the malleus and attached posteriorly to a cranial structure. The most feasible

cranial structure for attachment of the membrane is the posterior rim of the epitympanic recess where the tympanohyal is situated[17]. There are at least three reasons for this. First, such an attachment makes phylogenetic sense when compared with the condition in non-mammalian therapsids that have a squamosal sulcus[4]. Second, it is consistent with the condition of extant mammals in which the epitympanic recess is immediately dorsal or dorsomedial to the tympanic membrane[33–35]. Finally, developmental evidence shows that the membrane is supported by the tympanohyal and the element of Spence before full development of the ectotympanic[43]. This attachment is also consistent with the possibility that the posterior portion of the tympanic membrane is in contact with the malleus and incus, as evidenced by the fact that the lateral surfaces of the two ossicles are smooth and aligned even when the ossicles are in articulation.

The partly developed ectotympanic, absence of the manubrium of the malleus and lack of a fully suspended tympanic membrane in *Liaoconodon* echo the observations from developmental studies that phylogenetically older portions of the ossicles develop earlier than those that represent more recent evolutionary inventions[31,33]. Moreover, molecular–developmental studies also show that the manubrium, the processus brevis, the ectotympanic and tympanic membrane exhibit an interdependent relationship in the mammalian developmental programme[15,19]. Full development of the ectotympanic and the manubrium implies that they are probably neomorphs of mammals as co-members of an advanced complex adaptation late in therapsid evolution towards mammals[4,43].

The TMME

As the first unambiguous paleontological evidence, the middle ear of *Liaoconodon* corroborates the Reichert–Gaupp theory on the homology of the mammalian ear ossicles[7,8] as well as the hypothesis on the evolution of the mammalian middle ear[2,4]. The ear morphology of *Liaoconodon* represents a transitional stage in the evolution of mammalian middle ears regardless of how many times the DMME evolved[4,18,41]. The TMME can be characterized by several features: the articular, prearticular and angular lose their direct contact with the dentary (thus called as the malleus and ectotympanic) and are supported anteriorly by a persistent Meckel's cartilage, but not by cranial structures, in adult; the malleo-incudal articulation is hinge-like and lost its primary function for jaw suspension; all ear ossicles are primarily auditory structures but are not completely free from the feeding effect; the tympanic membrane is not fully suspended by the ectotympanic, and the manubrium of the malleus has not developed (Figs 2 and 3).

A key issue in the study of the evolution of the mammalian middle ear focuses on the detachment of the post-dentary bones[3,4,17,18,20], which primarily involves two events: separation of the ossicles from the dentary and degeneration of the Meckel's cartilage in adults[18]. The ear of *Liaoconodon* demonstrates that the transference from the mandibular middle ear to the TMME and then to the DMME involves complex morphological changes (Fig. 3 and Supplementary Information part C). Because of these changes, the TMME must be more efficient in airborne sound hearing than the mandibular middle ear, because the ossicles are proportionally smaller, the incus has more freedom and the other ossicles are separated from the dentary. In addition, the loose attachment of the OMC in the dentary, the sutured relationship of the OMC with the ossicles and the possible flexibility of the persistent Meckel's cartilage further enhance hearing by decoupling hearing from feeding. On the other hand, the hinged malleo-incudal joint, lack of the manubrium as a level arm in the ossicle chain and incomplete suspension of the tympanic membrane indicate that the middle ear of *Liaoconodon* is not so efficient as in extant mammals.

METHODS SUMMARY

Supplementary Information part F provides a full description of the methods.

1. Hopson, J. A. The origin of the mammalian middle ear. *Am. Zool.* **6,** 437–450 (1966).
2. Allin, E. F. Evolution of the mammalian middle ear. *J. Morphol.* **147,** 403–438 (1975).
3. Maier, W. Phylogeny and ontogeny of mammalian middle ear structures. *Neth. J. Zool.* **40,** 55–75 (1990).
4. Allin, E. F. & Hopson, J. A. in *The Evolutionary Biology of Hearing* (eds Webster, D. B., Fay, R. R. & Popper, A. N.) 587–614 (Springer, 1992).
5. Clack, J. A. & Allin, E. F. in *Evolution of the Vertebrate Auditory System* (eds Manley, G. A., Popper, A. N. & Fay, R. R.) 128–163 (Springer, 2004).
6. Lillegraven, J. A. & Krusat, G. Cranio-mandibular anatomy of *Haldanodon expectatus* (Docodonta; Mammalia) from the Late Jurassic of Portugal and its implications to the evolution of mammalian characters. *Contrib. Geol.* **28,** 39–138 (1991).
7. Reichert, C. Über die Visceralbogen der Wirbeltiere im Allgemeinen und deren Metamorphosen bei den Vögeln und Säugetieren. *Arch. Anat. Physiol. Medizin* **1837,** 120–222 (1837).
8. Gaupp, E. Die Reichertsche Theorie (Hammer-, Amboss- und Kieferfrage). *Arch. Anat. Entwick.* **1912,** 1–416 (1913).
9. Goodrich, E. S. *Studies on the Structure and Development of Vertebrates* (Macmillan, 1930).
10. Kermack, K. A., Mussett, F. & Rigney, H. W. The lower jaw of *Morganucodon. Zool. J. Linn. Soc.* **53,** 87–175 (1973).
11. Kermack, K. A., Mussett, F. & Rigney, H. W. The skull of *Morganucodon. Zool. J. Linn. Soc.* **71,** 1–158 (1981).
12. Kielan-Jaworowska, Z., Cifelli, R. L. & Luo, Z.-X. *Mammals from the Age of Dinosaurs— Origins, Evolution, and Structure* (Columbia Univ. Press, 2004).
13. Zeller, U. Die Entwicklung und Morphologie des Schädels von *Ornithorhynchus anatinus* (Mammalia: Prototheria: Monotremata). *Abhandl. Senckenberg. Natur. Gesell.* **545,** 1–188 (1989).
14. Zeller, U. in *Mammal Phylogeny* Vol. 1 (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 95–107 (Springer, 1993).
15. Mallo, M. Formation of the middle ear: recent progress on the developmental and molecular mechanisms. *Dev. Biol.* **231,** 410–419 (2001).
16. Wang, Y.-Q., Hu, Y.-M., Meng, J. & Li, C.-K. An ossified Meckel's cartilage in two Cretaceous mammals and origin of the mammalian middle ear. *Science* **294,** 357–361 (2001).
17. Meng, J. Hu, Y.-M., Wang, Y.-Q. & Li, C.-K. The ossified Meckel's cartilage and internal groove in Mesozoic mammaliaforms: implications to origin of the definitive mammalian middle ear. *Zool. J. Linn. Soc.* **138,** 431–448 (2003).
18. Luo, Z.-X., Chen, P.-J., Li, G. & Chen, M. A new eutriconodont mammal and evolutionary development in early mammals. *Nature* **446,** 288–293 (2007).
19. Takechi, M. & Kuratani, S. History of studies on mammalian middle ear evolution: a comparative morphological and developmental biology perspective. *J. Exp. Zool. B* **314,** 1–17 (2010).
20. Rowe, T. B. Coevolution of the mammalian middle ear and neocortex. *Science* **273,** 651–654 (1996).
21. He, H.-Y. *et al.* Timing of the Jiufotang Formation (Jehol Group) in Liaoning, northeastern China, and its implications. *Geophys. Res. Lett.* **31,** L12605 (2004).
22. Ji, Q., Luo, Z.-X. & Ji, S.-A. A Chinese triconodont mammal and mosaic evolution of the mammalian skeleton. *Nature* **398,** 326–330 (1999).
23. Li Li, C.-K., Hu, Y.-M., Wang, Y.-Q. & Meng, J. A new species of *Gobiconodon* (Triconodonta, Mammalia) and its implication for the age of Jehol Biota. *Chin. Sci. Bull.* **48,** 1129–1134 (2003).
24. Meng, J., Hu, Y.-M., Wang, Y.-Q. & Li, C.-K. A new gobiconodont species (Mammalia) from the Early Cretaceous Yixian Formation of Liaoning, China. *Vert. PalAsiatica* **43,** 1–10 (2005).
25. Li, J.-L., Wang, Y., Wang, Y.-Q. & Li, C. K. A new family of primitive mammal from the Mesozoic of western Liaoning, China. *Chin. Sci. Bull.* **45,** 2545–2549 (2000).
26. Hu, Y.-M., Meng, J., Li, C.-K. & Wang, Y.-Q. Large Mesozoic mammals fed on young dinosaurs. *Nature* **433,** 149–153 (2005).
27. Ji, Q., Luo, Z.-X., Zhang, X., Yuan, C.-X. & Xu, L. Evolutionary development of the middle ear in Mesozoic therian mammals. *Science* **326,** 278–281 (2009).
28. McClain, J. A. The development of the auditory ossicles of the opossum (*Didelphys virginiana*). *J. Morphol.* **64,** 211–265 (1939).
29. Filan, S. L. Development of the middle ear region in *Monodelphis domestica* (Marsupialia, Didelphidae): marsupial solutions to an early birth. *J. Zool.* **225,** 577–588 (1991).
30. Clark, C. T. & Smith, K. K. Cranial osteogenesis in *Monodelphis domestica* (Didelphidae) and *Macropus eugenii* (Macropodidae). *J. Morphol.* **215,** 119–145 (1993).
31. Sánchez-Villagra, M. R., Gemballa, S., Nummela, S., Smith, K. K. & Maier, W. Ontogenetic and phylogenetic transformations of the ear ossicles in marsupial mammals. *J. Morphol.* **251,** 219–238 (2002).
32. Meng, J. & Wyss, A. R. Monotreme affinities and low-frequency hearing suggested by multituberculate ear. *Nature* **377,** 141–144 (1995).
33. Fleischer, G. Studien am Skelett des Gehörorgans der Säugetiere, einschliesslich des Menschen. *Säugetierk. Mitt.* **21,** 131–239 (1973).
34. Fleischer, G. Evolutionary principles of the mammalian middle ear. *Adv. Anat. Embryol. Cell Biol.* **55,** 1–70 (1978).
35. Henson, O. W. Comparative anatomy of the middle ear. *Handbook of Sensory Physiology: The Auditory System* Vol. 5, Pt 1 (eds Keidel, W. D. & Neff, W. D.) 39–110 (Springer, 1974).
36. Barghusen, H. R. in *The Ecology and Biology of Mammalian-like Reptiles* (eds Hotton III N., MacLean P. D., Roth, J. J. & Roth, E. C.) 253–262 (Smithsonian Institution Press, 1986).
37. Crompton, A. W. in *Studies in Vertebrate Evolution* (eds Joysey, K. A. & Kemp, T. S.) 231–251 (Oliver & Boyd, 1972).
38. Bensley, B. A. On the identification of Meckelian and mylohyoid grooves in the jaws of Mesozoic and Recent Mammalia. *Univ. Toronto Stud. Biol. Ser.* **3,** 75–81 (1902).
39. Krebs, B. Evolution of the mandible and lower dentition in dryolestids (Pantotheria, Mammalia). *Zool. J. Linn. Soc.* **50** (Suppl. 1), 89–102 (1971).
40. Rougier, G. W., Wible, J. R. & Novacek, M. J. Middle-ear ossicles of the multituberculate *Kryptobaatar* from the Mongolian Late Cretaceous: implications for mammaliamorph relationships and the evolution of the auditory apparatus. *Am. Mus. Novit.* **3187,** 1–43 (1996).
41. Allin, E. F. in *The Ecology and Biology of Mammal-like Reptiles* (eds Hotton, N., MacLean P. D., Roth, J. J. & Roth, E. C.) 283–294 (Smithsonian Institution Press, 1986).
42. Kermack, K. A. & Mussett, F. The ear in mammal-like reptiles and early mammals. *Acta Palaeontol. Pol.* **28,** 147–158 (1983).
43. Presley, R. Lizards, mammals and the primitive tetrapod tympanic membrane. *Symp. Zool. Soc. Lond.* **52,** 127–152 (1984).
44. Westoll, T. S. The mammalian middle ear. *Nature* **155,** 114–115 (1945).
45. Watson, D. M. S. Evolution of the mammalian ear. *Evolution* **7,** 159–177 (1953).
46. Shute, C. C. The evolution of the mammalian eardrum and tympanic cavity. *J. Anat.* **90,** 261–281 (1956).

ARTICLE

Loss-of-function mutations in sodium channel Na$_v$1.7 cause anosmia

Jan Weiss[1]*, Martina Pyrski[1]*, Eric Jacobi[1], Bernd Bufe[1], Vivienne Willnecker[2], Bernhard Schick[2], Philippe Zizzari[3], Samuel J. Gossage[4], Charles A. Greer[5], Trese Leinders-Zufall[1], C. Geoffrey Woods[6], John N. Wood[4,7] & Frank Zufall[1]

Loss of function of the gene *SCN9A*, encoding the voltage-gated sodium channel Na$_v$1.7, causes a congenital inability to experience pain in humans. Here we show that Na$_v$1.7 is not only necessary for pain sensation but is also an essential requirement for odour perception in both mice and humans. We examined human patients with loss-of-function mutations in *SCN9A* and show that they are unable to sense odours. To establish the essential role of Na$_v$1.7 in odour perception, we generated conditional null mice in which Na$_v$1.7 was removed from all olfactory sensory neurons. In the absence of Na$_v$1.7, these neurons still produce odour-evoked action potentials but fail to initiate synaptic signalling from their axon terminals at the first synapse in the olfactory system. The mutant mice no longer display vital, odour-guided behaviours such as innate odour recognition and avoidance, short-term odour learning, and maternal pup retrieval. Our study creates a mouse model of congenital general anosmia and provides new strategies to explore the genetic basis of the human sense of smell.

The inability to sense odours is known as general anosmia; individuals born with this phenotype are afflicted with congenital general anosmia. Except for some syndromic cases such as Kallmann syndrome, no causative genes for human congenital general anosmia have been identified so far[1–3]. Nine mammalian genes encoding voltage-gated sodium channel α-subunits have been cloned and shown to be differentially expressed in the nervous system[4,5]. Of these, *SCN9A*, encoding the tetrodotoxin (TTX)-sensitive sodium channel Na$_v$1.7, has received specific attention because of its key role in human pain perception. Individuals carrying loss-of-function mutations in *SCN9A* are unable to experience pain, and an essential requirement of Na$_v$1.7 function for nociception in humans has been established[6–9]. Whether all other sensory modalities are fully preserved in these individuals remained unclear, although an association between congenital inability to experience pain and sense of smell deficits has been suggested[7]. In this study we examine human patients carrying *SCN9A* loss-of-function mutations and demonstrate that they fail to sense odours. We establish a mouse model of congenital general anosmia and provide mechanistic insight into the role of Na$_v$1.7 in olfaction. Together with previous findings[6–8,10], our results establish that loss-of-function mutations in a single gene, *SCN9A*, cause a general loss of two major senses—nociception and smell—thus providing a mechanistic link between these two sensory modalities.

Requirement for Na$_v$1.7 in human olfaction

Three individuals with congenital analgesia were ascertained and studied. All three were in their third decade of life and had never experienced acute pain but had no other neurological, cognitive, growth, appearance or health problems. All had suffered from multiple painless fractures and other injuries. Two had given birth painlessly. A working diagnosis of channelopathy-associated insensitivity to pain (CAIP) was made and in each *SCN9A* was sequenced[6]. In the first,

who has been the subject of a detailed case report, the mutations c.774_775delGT and c.2488C>T were found[10]. These mutations, frameshift and nonsense, respectively, would be predicted to lead to a lack of functional Na$_v$1.7 protein. The other two were siblings and had the mutations c.4975A>T and c.3703delATAGCATATGG; again, nonsense and frameshift mutations and predicted to lead to no functional Na$_v$1.7 protein. The mother of the siblings was found to be heterozygous for the 11-base-pair deletion and the father heterozygous for the nonsense mutation. Therefore the diagnosis of CAIP was substantiated. We next assessed their sense of smell; none complained of having no sense of smell, one had been a cigarette smoker, none had chronic nasal problems. In the first woman smell function was assessed by using the University of Pennsylvania Smell Identification Test (UPSIT), a standardized 40-item smell test. The results revealed that she was unable to detect any of the odours (Fig. 1a, black bar). Nine healthy, young individuals served as controls (Fig. 1a, grey bars). In the sibling pair we assessed the parents and their two affected offspring together. All were tested in sequence with cotton wool pads suffused with selected odour stimuli: balsamic vinegar, orange, mint, perfume, water (control) and coffee. Both parents correctly identified all stimuli, including smelling nothing for the water. The siblings detected none of the odours. For the siblings the test was repeated using subjectively unpleasant amounts of balsamic vinegar and perfume: the parents identified the odours correctly and found them unpleasant; the siblings neither identified the odours nor experienced any discomfort.

We proposed that these odour-sensing deficits are caused by loss of Na$_v$1.7 function in olfactory sensory neurons (OSNs). Indeed, when we investigated expression of Na$_v$1.7 in normal human olfactory epithelium, we detected messenger RNA for Na$_v$1.7 and the GTP-binding protein Gα$_{olf}$, a prototypical signature of classical OSNs (Fig. 1b). Immunohistochemistry using an antibody specific to Na$_v$1.7 verified that Na$_v$1.7 is normally expressed in human OSNs (Fig. 1c, d).
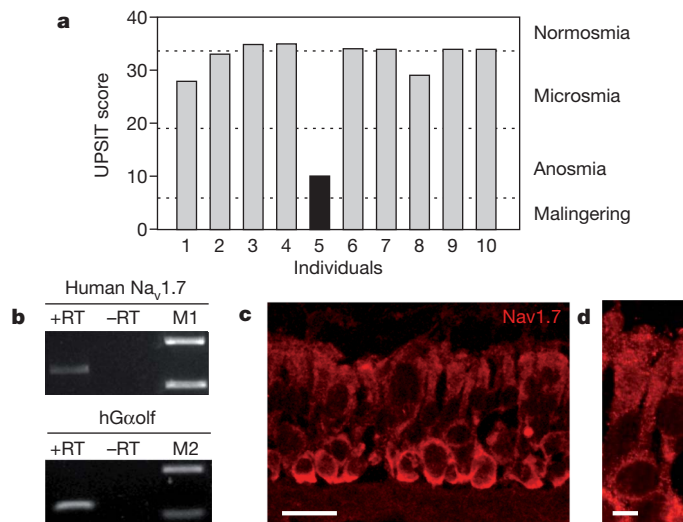
Figure 1 | Na$_v$1.7 in human olfaction. a, Quantified olfactory assessment of the first individual with confirmed Na$_v$1.7 loss-of-function mutations (black bar) using the standardized, 40-item UPSIT test showed that she was unable to detect any of the odour stimuli; the test score revealed general anosmia in this patient. Nine healthy, young individuals served as controls (grey bars). We assessed odour perception in two other individuals with confirmed Na$_v$1.7 loss-of-function mutations and both were unable to sense any of the odours. These results are described in the main text. **b**, Expression of Na$_v$1.7 in olfactory epithelium from unaffected normal humans. RT–PCR products with gene-specific primers for human Na$_v$1.7 (top; size, 1,128 bp) and the G-protein Gα$_{olf}$ (bottom; size, 143 bp). PCRs were carried out with equal amounts of RNA in the presence (+RT) or absence (−RT) of reverse transcriptase to exclude product amplification from genomic DNA. M1, size marker; 2,000 bp, 850 bp; M2, size marker; 400 bp, 100 bp. Similar results were obtained in two other human olfactory mucosa samples. **c**, Confocal fluorescence image of Na$_v$1.7 immunoreactivity (red) in a cryosection of human olfactory epithelium. Scale bar, 20 μm. **d**, Enlargement showing a single OSN. Scale bar, 5 μm.

Conditional Na$_v$1.7 null mice

To investigate the mechanisms that underlie the essential role of Na$_v$1.7 in odour perception, we first examined Na$_v$1.7 expression in the mouse olfactory system and then used the Cre-*loxP* system to delete the channel in those cells that express olfactory marker protein (OMP), which includes all classical OSNs[11]. These mice enabled us to examine the mechanisms underlying Na$_v$1.7-associated anosmia and the behavioural consequences. Consistent with our findings in human OSNs, OSNs from wild-type mice (C57BL/6, referred to as B6) showed Na$_v$1.7 immunoreactivity at their somata (Supplementary Fig. 1). Of greater interest, coronal sections containing main olfactory epithelium (MOE), olfactory nerves and the two olfactory bulbs revealed the most marked Na$_v$1.7 staining in the superficial olfactory nerve layer (ONL, containing axons from OSNs) as well as the glomerular layer (a complex neuropil that includes the presynaptic OSN boutons) of the olfactory bulb (Fig. 2a–c). Higher magnification of individual glomeruli verified co-expression of Na$_v$1.7 with OMP in the glomerular neuropil (Fig. 2b), whereas olfactory bulb projection neurons (the mitral/tufted or M/T cells) and local interneurons did not show Na$_v$1.7 immunoreactivity (Fig. 2a). Thus, Na$_v$1.7 occupies a critical presynaptic location at the first synapse in the olfactory system.

Na$_v$1.7 is not the sole Na$_v$ channel expressed in mouse OSNs. Real-time quantitative polymerase chain reaction with reverse transcription (qRT–PCR) analysis identified Na$_v$1.3 as an additional candidate (Supplementary Fig. 2) and immunohistochemistry verified its expression in OSNs and their axons (Fig. 2d). However, unlike Na$_v$1.7 we did not observe Na$_v$1.3 immunoreactivity in individual glomeruli (Fig. 2d), indicating that Na$_v$1.7 could be the sole Na$_v$ channel underlying action potential propagation in olfactory glomeruli and OSN nerve terminals.

To create a conditional knockout mouse model, we crossed 'floxed' Na$_v$1.7 mice harbouring a *loxP*-flanked *Scn9a* gene[12] to homozygous OMP–Cre mice in which the OMP-coding region is replaced by that of Cre recombinase[13]. Further breeding established offspring that were both homozygous for the floxed *Scn9a* alleles and heterozygous for *cre* and *Omp*. In these mice, Cre-mediated Na$_v$1.7 deletion was restricted to OMP-positive cells (henceforward referred to as cNa$_v$1.7$^{-/-}$ mice). These mice lacked Na$_v$1.7 expression in a tissue-specific manner
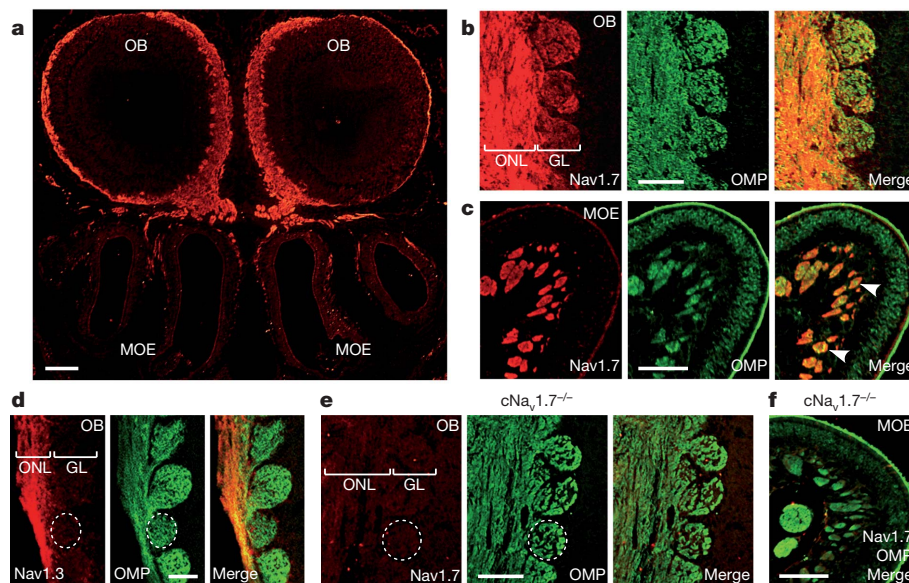
Figure 2 | Na$_v$1.7 expression in the mouse main olfactory system. a, Na$_v$1.7 immunoreactivity (red) detected in the olfactory bulbs (OB, top) and the underlying main olfactory epithelium (MOE, bottom). B6 mouse, 2-days old. **b**, Colocalization of Na$_v$1.7 (red) and OMP (green) expression in the ONL and glomerular layer (GL) of the olfactory bulb. B6 mouse, 3-weeks old. **c**, Strong Na$_v$1.7 immunoreactivity in OSN axon bundles (arrowheads). B6 mouse, 3-weeks old. **d**, Na$_v$1.3 expression (red) terminates in the ONL and is not detectable in OMP-labelled (green) glomeruli. OMP–GFP mouse, 16-days old. **e–f**, cNa$_v$1.7$^{-/-}$ mice lack Na$_v$1.7 immunoreactivity in the ONL and glomerular layer of the olfactory bulb (**e**) as well as in OSN axon bundles of MOE (**f**). Dashed circle, individual glomerulus. Scale bars: **a**, 200 μm; **b, c, e, f**, 100 μm; **d**, 50 μm.

(Fig. 2e, f and Supplementary Fig. 3). Successful matings occurred between $cNa_v1.7^{+/-}$ males and $cNa_v1.7^{-/-}$ females whereas homozygous knockout pairs did not produce any offspring. $cNa_v1.7^{-/-}$ mice showed a reduced body weight during the first three months of postnatal development (Supplementary Fig. 4a). Because both $Na_v1.7$ (refs 14, 15) and OMP[16] are also expressed in some neurons mediating hormonal regulation, we assayed insulin-like growth factor (IGF-1, also known as somatomedin C) in $cNa_v1.7^{-/-}$ (650 ± 94 ng ml^{-1}; $n = 4$) versus $cNa_v1.7^{+/-}$ mice (684 ± 27 ng ml^{-1}; $n = 4$; mean ± s.d.) but found no significant difference between the two genotypes ($P = 0.26$). Given that newborn $cNa_v1.7^{-/-}$ mice had very little milk in their stomachs (Supplementary Fig. 4b), the diminished weight gain was probably caused by a deficit to suckle effectively, consistent with results in mice deficient in $G\alpha_{olf}$ (*Gnal*) or the cAMP-gated cation channel (*Cnga2*)[17,18].

Loss of synaptic transfer in olfactory glomeruli

To define the function of $Na_v1.7$ in OSNs, we prepared MOE tissue slices[19] and recorded sodium currents in voltage-clamped OSNs. Both $Na_v1.7^{+/-}$ and $Na_v1.7^{-/-}$ OSNs displayed sizeable, TTX-sensitive sodium currents in response to step depolarizations (Fig. 3a, b). On the basis of its biophysical properties, $Na_v1.7$ has been suggested to transduce generator potentials into action potentials in sensory neurons[9]. However, peak current densities of voltage-activated sodium currents were reduced only moderately, by about 20%, in $Na_v1.7^{-/-}$ OSNs (Fig. 3b). To determine whether $Na_v1.7^{-/-}$ OSNs could still produce odour-evoked action potentials, we used extracellular loose-patch recording from visually identified OSN dendritic knobs[20] and analysed spike frequency histograms after brief odour exposure (Fig. 3c). There was no obvious difference in odour responsiveness in $Na_v1.7^{-/-}$ versus $Na_v1.7^{+/-}$ OSNs (Fig. 3c). We obtained similar results when we stimulated the cells with 3-isobutyl-1-methylxanthine (IBMX)[21], which raises intracellular cAMP by inhibiting endogenous phosphodiesterase activity (Fig. 3c). Thus, although the initial site of odour-evoked action potential generation in OSNs is unknown, $Na_v1.7$ is not essential for this activity.

Because $Na_v1.7$ is expressed in olfactory bulb glomeruli (Fig. 2), we reasoned that it could be required for action potential conduction in OSN terminals. Olfactory glomeruli are delineated spheres of neuropil containing synapses from the OSN axon terminals onto juxtaglomerular interneurons and M/T projection neurons[22,23]. To examine whether presynaptic activity of $Na_v1.7$ underlies transmitter release in the olfactory glomerulus, we prepared olfactory bulb tissue slices[24] and combined ONL focal electric stimulation with whole-cell patch-clamp recording from visually identified M/T cells. With the chosen protocol, in control $cNa_v1.7^{+/-}$ mice a single electrical stimulus in the ONL produced a reliable postsynaptic response in M/T cells. Under current clamp, such responses consisted of a prolonged excitation lasting on average for 2.4 ± 0.4 s (Fig. 3d, top; $n = 29$), with response latencies of 22 ± 4 ms ($n = 29$). Under voltage clamp, we observed bursts of postsynaptic currents (Fig. 3f; duration, 3.2 ± 0.4 s; $n = 26$). In stark contrast, in the $cNa_v1.7^{-/-}$ mice such postsynaptic responses were virtually absent in M/T cells, even when the stimulus strength was increased by several-fold (Fig. 3d–f; $n = 49$). Importantly, M/T cells in these mice still produced normal action potentials when depolarized via current injection through the patch pipette (Fig. 3d, bottom), consistent with the fact that M/T cells lack both OMP and $Na_v1.7$ expression (Fig. 2) and indicating that the effect of deleting $Na_v1.7$ is presynaptic to the M/T cells. The inability of M/T cells to produce synaptic responses to ONL stimulation was not due to a potential deficit in synapse formation because: (1) immunohistochemistry showed normal expression of the vesicular glutamate transporter 2 (vGluT2, which is selectively expressed in OSN axon terminals)[25,26] (Supplementary Fig. 5); and (2) electron microscopy revealed the existence of normal OSN boutons and synapses in the glomeruli of $cNa_v1.7^{-/-}$ mice (Supplementary Fig. 6). Furthermore, conditional
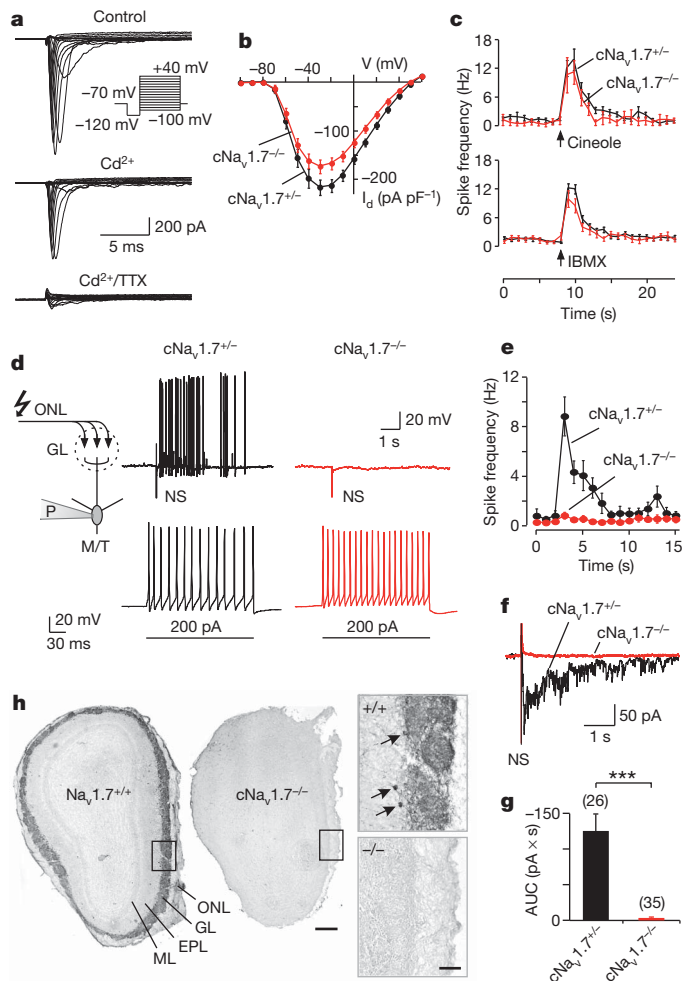
Figure 3 | $Na_v1.7$ is essential for synaptic transfer in the olfactory glomerulus. a, Families of whole-cell currents to a series of depolarizing voltage steps recorded from $cNa_v1.7^{-/-}$ OSNs. OSNs were exposed successively to extracellular bath solution (control), bath solution containing Cd^{2+} (100 μM), and bath solution containing Cd^{2+} (100 μM) and TTX (2 μM). Holding potential, −70 mV. **b**, Current-density–voltage curves of sodium currents from $cNa_v1.7^{+/-}$ (black; $n = 35$) and $cNa_v1.7^{-/-}$ OSNs (red; $n = 46$). Current densities of $Na_v1.7^{+/-}$ OSNs were significantly diminished between −30 mV and 10 mV (LSD, $P = 0.001–0.04$). **c**, Action potential responses in visually identified OSN dendritic knobs to pulsed stimulation (0.2 s) with cineole (100 μM) or IBMX (100 μM) in $cNa_v1.7^{+/-}$ (black; $n = 28$ and 78 cells, respectively) and $cNa_v1.7^{-/-}$ OSNs (black; $n = 5$ and 25, respectively). Firing properties were similar in both genotypes (LSD, $P = 0.14–0.73$). **d**, M/T cells exhibit postsynaptic potentials to presynaptic nerve stimulation (NS) in $cNa_v1.7^{+/-}$ (black), but not in $cNa_v1.7^{-/-}$ mice (red), whereas M/T cells in both genotypes show normal action potentials to current injection (200 pA). Current-clamp whole-cell recording. **e**, Absence of M/T cell postsynaptic potentials to nerve stimulation in $cNa_v1.7^{-/-}$ (red; $n = 44$) versus $cNa_v1.7^{+/-}$ (black; $n = 29$) (LSD, $P < 0.01–0.001$). **f**, Example of postsynaptic currents in M/T cells of $cNa_v1.7^{+/-}$ (black) and $cNa_v1.7^{-/-}$ mice (red). **g**, Analysis of area under curve (AUC) of M/T cell postsynaptic currents during a 5 s interval after nerve stimulation in $cNa_v1.7^{+/-}$ (black) and $cNa_v1.7^{-/-}$ mice (red). Number of cells tested is shown in brackets above each bar. Unpaired t-test: ***$P < 0.0001$. **h**, Marked reduction of TH expression in juxtaglomerular cells of the olfactory bulb in $cNa_v1.7^{-/-}$ mice. EPL, external plexiform layer; ML, mitral cell layer. Arrows in the inset (+/+) indicate individual juxtaglomerular cells. Scale bars: overview, 100 μm; inset, 50 μm. Error bars represent mean ± s.e.m.

OSN expression of tetanus toxin light chain, which inhibits synaptic release, does not alter the pattern of axonal targeting in olfactory bulb glomeruli during development[27].

Tyrosine hydroxylase (TH) expression in juxtaglomerular neurons of the olfactory bulb, a correlate of afferent trans-synaptic activity,

requires olfactory nerve input and odour-stimulated glutamate release by OSN terminals[28]. Consistent with a loss of OSN synaptic release, TH expression was markedly reduced in cNa$_v$1.7$^{-/-}$ mice (Fig. 3h; n = 6). The level of TH downregulation was similar to that observed after odour deprivation by naris occlusion[29] or after deletion of the *Cnga2* cation channel gene[30]. Thus, we conclude that the presence of Na$_v$1.7 in OSN axons is an essential and non-redundant requirement to initiate information transfer from OSN terminals to neurons in the olfactory bulb.

The absence of odour–guided behaviours

To further validate these results, we investigated several odour-guided behaviours in B6, cNa$_v$1.7$^{+/-}$ and cNa$_v$1.7$^{-/-}$ mice. First, we performed an odour preference test[31] to assess recognition abilities for innate odour qualities (Fig. 4a). Filter papers scented with various cues representing both species-specific and food odours (male and female urine, peanut butter, milk) were presented to the mice and investigation times were analysed. Water was used as a neutral stimulus and 1,8-cineole (eucalyptol), which does not evoke innate attraction, served as the control (n ≥ 7 for each cue and strain, respectively). B6 and cNa$_v$1.7$^{+/-}$ mice both showed strong attraction towards conspecific and food odours, whereas cNa$_v$1.7$^{-/-}$ mice failed to show any interest in these stimuli.

Second, we explored whether Na$_v$1.7 is required for innate avoidance behaviour towards a predator odour, trimethyl-thiazoline (TMT)[31], which is normally secreted from the fox anal gland and known to induce aversive behaviour and fear responses in mice. We observed robust avoidance behaviour in both B6 (n = 6) and cNa$_v$1.7$^{+/-}$ mice (n = 5) but, notably, cNa$_v$1.7$^{-/-}$ mice lacked an innately aversive response in this assay (n = 5; Fig. 4b, c).

Third, we investigated the performance of cNa$_v$1.7$^{-/-}$ mice in a habituation–dishabituation assay, which allows for measurement of novel odour investigation, short-term odour learning, and odour discrimination[32] (Fig. 4d). Mice of both sexes were each presented three distinct stimuli (water, female urine, male urine), each delivered for three successive trials, and investigation time during each trial (3 min) was analysed. Consistent with the results of Fig. 4a, cNa$_v$1.7$^{-/-}$ mice (n = 8) failed to show significant odour investigation, habituation, or discrimination abilities when compared with B6 (n = 8) or cNa$_v$1.7$^{+/-}$ mice (n = 8) (Fig. 4d; least significant difference (LSD), P < 0.0001).

Last, we examined pup retrieval ability of female mice, a social behaviour that probably depends on a functional main olfactory system (Fig. 4e). Three pups of a litter were removed from the nest, randomly distributed in the cage, and the time to retrieve each pup into the nest was quantified. In contrast to the performance of B6 (n = 12) or cNa$_v$1.7$^{+/-}$ mice (n = 6), cNa$_v$1.7$^{-/-}$ mice (n = 5) failed to retrieve any of the three pups during a 10-min trial period (Fig. 4e).

Conclusions and prospects

Our results establish a critical role of the Na$_v$1.7 sodium channel in olfaction. Using conditional Na$_v$1.7 null mice, we demonstrate that, in the absence of Na$_v$1.7, OSNs are still electrically active and generate odour-evoked action potentials but fail to initiate synaptic signalling to the projection neurons in the olfactory bulb. These results provide evidence that Na$_v$1.7 is an essential and non-redundant requirement for action potential propagation in the sections of OSN axons within the olfactory glomerulus. The conditional null mice no longer show a wide range of vital, odour-guided behaviours including innate attraction to food and conspecific odours, odour discrimination and short-term odour learning, innate avoidance towards a predator odour, effective suckling behaviour of newborn pups, and maternal pup retrieval. Within the limits of our anatomical analyses, synapse formation in these mice appears normal, indicating that the behavioural phenotype of the mutant mice is most likely the result of a loss of signalling at the first synapse in the olfactory system. Whether Na$_v$1.7 or other sodium channel subunits such as Na$_v$1.3 are involved in OSN
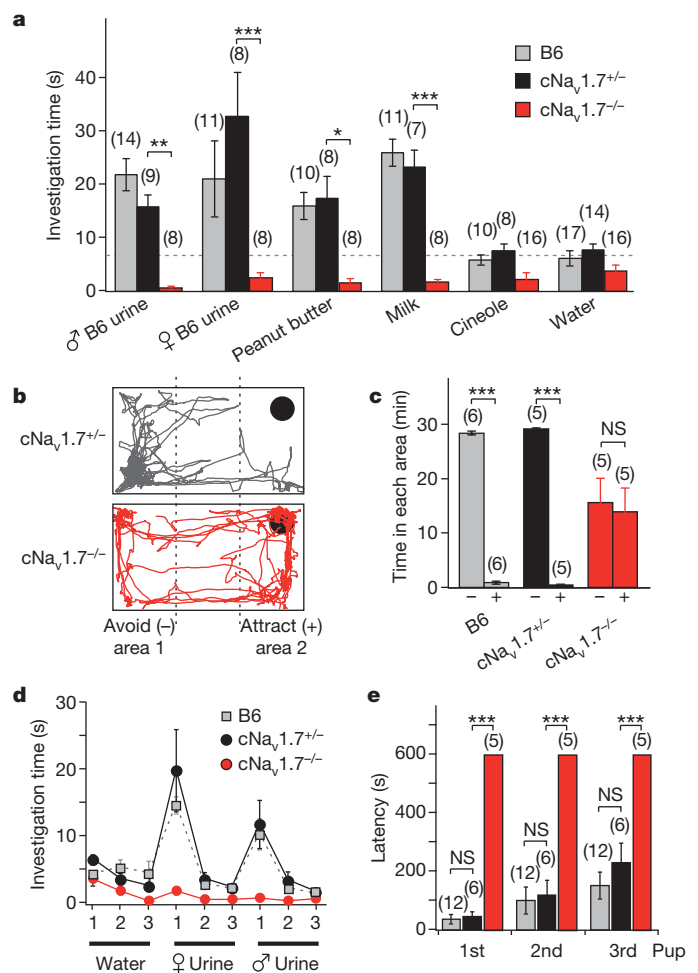
Figure 4 | cNa$_v$1.7$^{-/-}$ mice are anosmic. a, Innate olfactory preference test. Mean investigation times were quantified in B6, cNa$_v$1.7$^{+/-}$ and cNa$_v$1.7$^{-/-}$ mice during a 3-min test period. Mean investigation time for water in B6 and cNa$_v$1.7$^{+/-}$ mice (dashed line) served as a marker for attraction. Number of animals tested is shown in brackets above each bar. LSD: *P < 0.02; **P < 0.001; ***P < 0.0001. No difference was observed between B6 and cNa$_v$1.7$^{+/-}$ mice (LSD, P = 0.1). **b, c,** Innate olfactory avoidance to the predator odour TMT. **b,** Examples of the trajectory plot of the position of a cNa$_v$1.7$^{+/-}$ and a cNa$_v$1.7$^{-/-}$ mouse (30-min video-tracking). Location of TMT (5 µl) indicated by the circle in the upper right corner. **c,** Quantification of time spent in either area 1 (−) or area 2 (+) during exposure to TMT indicates that cNa$_v$1.7$^{-/-}$ mice lack avoidance behaviour to TMT (LSD, P = 0.67). LSD: ***P < 0.0001. **d,** Olfactory habituation–dishabituation assay. Mean investigation time of B6, cNa$_v$1.7$^{+/-}$ and cNa$_v$1.7$^{-/-}$ mice (n = 8 for each group) to three distinct stimuli (water, female and male mouse urine) during a 3-min test period were quantified. Numbers indicate stimulus presentation order. **e,** Pup retrieval test. Mean latencies of B6, cNa$_v$1.7$^{+/-}$ and cNa$_v$1.7$^{-/-}$ female mice in retrieving three individual pups that were randomly distributed throughout the cage. cNa$_v$1.7$^{-/-}$ mice failed to retrieve any of the pups during a 10-min trial period (LSD, P < 0.0001). NS, not significant (LSD, P = 0.41). Error bars represent mean ± s.e.m.

axon pathfinding and activity-dependent neural map formation[33] in the mouse olfactory system remains to be seen. Importantly, the phenotype of the mutant mice—the inability to perceive odours—is similar to that observed in human patients with confirmed Na$_v$1.7 loss-of-function mutations. Smell tests in three individuals with congenital analgesia establish that they are unable to sense any of the odours. Systematic olfactory testing of patients carrying Na$_v$1.7 loss-of-function mutations will be required in the future.

The genetic basis of sensory deficits such as blindness, deafness and pain disorders has been extensively studied in recent years. By comparison, relatively little progress has been made in understanding

human congenital general anosmia[1]. Mutations in olfactory signal transduction genes such as *CNGA2*, *GNAL* and *ADCY3* do not seem to be a major cause of human congenital general anosmia[2]. The identification of a sodium channel subunit as a causative gene for an inherited form of general anosmia provides new insight into the molecular pathophysiology of olfaction and should stimulate further research aimed at understanding the genetic basis of the human sense of smell.

METHODS SUMMARY

Human subjects. All research involving humans was obtained with the informed consent of the patients and performed under protocols approved by the Ethics Committee of the relevant institution. Human nasal mucosa was obtained by biopsy during routine nasal surgery. Further details of the human smell tests can be found in Methods.

Animals. The relevant Institutional Animal Care and Use Committee approved all procedures. Experiments used tissue-specific $Na_v1.7$-deficient, C57BL/6J and OMP–GFP mice. See Methods for details.

PCR analyses. Human olfactory mucosa samples were examined individually whereas mouse tissue was pooled from four different animals. PCR products were amplified with gene-specific primers and specificity was controlled by sequencing. Primers are specified in Methods.

Immunohistochemistry and electron microscopy. These followed previously published procedures[26,34,35] as described in Methods.

Electrophysiology. We used intact mouse MOE preparations as described previously[19,20]. Details of the electrophysiological experiments[36] are given in Methods.

Behavioural analyses. Innate olfactory preference tests and avoidance measurements used methods similar to those previously described[31,37]. Olfactory habituation–dishabituation followed previously described procedures[32]. Details are given in Methods.

IGF-1 assays. These were done as described in Methods.

Statistics. Details of the statistical tests are given in Methods. Unless otherwise stated, results are presented as means ± s.e.m.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Hasin-Brumshtein, Y., Lancet, D. & Olender, T. Human olfaction: from genomic variation to phenotypic diversity. *Trends Genet.* **25,** 178–184 (2009).
2. Feldmesser, E. *et al.* Mutations in olfactory signal transduction genes are not a major cause of human congenital general anosmia. *Chem. Senses* **32,** 21–30 (2007).
3. Keller, A. & Vosshall, L. B. Better smelling through genetics: mammalian odor perception. *Curr. Opin. Neurobiol.* **18,** 364–369 (2008).
4. Goldin, A. L. Resurgence of sodium channel research. *Annu. Rev. Physiol.* **63,** 871–894 (2001).
5. Catterall, W. A., Goldin, A. L. & Waxman, S. G. International Union of Pharmacology. XLVII. Nomenclature and structure–function relationships of voltage-gated sodium channels. *Pharmacol. Rev.* **57,** 397–409 (2005).
6. Cox, J. J. *et al.* An *SCN9A* channelopathy causes congenital inability to experience pain. *Nature* **444,** 894–898 (2006).
7. Goldberg, Y. P. *et al.* Loss-of-function mutations in the $Na_v1.7$ gene underlie congenital indifference to pain in multiple human populations. *Clin. Genet.* **71,** 311–319 (2007).
8. Ahmad, S. *et al.* A stop codon mutation in SCN9A causes lack of pain sensation. *Hum. Mol. Genet.* **16,** 2114–2121 (2007).
9. Dib-Hajj, S. D., Cummins, T. R., Black, J. A. & Waxman, S. G. From genes to pain: $Na_v1.7$ and human pain disorders. *Trends Neurosci.* **30,** 555–563 (2007).
10. Nilsen, K. B. *et al.* Corrigendum to "Two novel SCN9A mutations causing insensitivity to pain." *Pain* **145,** 264 (2009).
11. Munger, S. D., Leinders-Zufall, T. & Zufall, F. Subsystem organization of the mammalian sense of smell. *Annu. Rev. Physiol.* **71,** 115–140 (2009).
12. Nassar, M. A. *et al.* Nociceptor-specific gene deletion reveals a major role for $Na_v1.7$ (PN1) in acute and inflammatory pain. *Proc. Natl Acad. Sci. USA* **101,** 12706–12711 (2004).
13. Li, J., Ishii, T., Feinstein, P. & Mombaerts, P. Odorant receptor gene choice is reset by nuclear transfer from mouse olfactory sensory neurons. *Nature* **428,** 393–399 (2004).
14. Klugbauer, N., Lacinova, L., Flockerzi, V. & Hofmann, F. Structure and functional expression of a new member of the tetrodotoxin-sensitive voltage-activated sodium channel family from human neuroendocrine cells. *EMBO J.* **14,** 1084–1090 (1995).
15. Morinville, A. *et al.* Distribution of the voltage-gated sodium channel $Na_v1.7$ in the rat: expression in the autonomic and endocrine systems. *J. Comp. Neurol.* **504,** 680–689 (2007).
16. Baker, H., Grillo, M. & Margolis, F. L. Biochemical and immunocytochemical characterization of olfactory marker protein in the rodent central nervous system. *J. Comp. Neurol.* **285,** 246–261 (1989).
17. Belluscio, L., Gold, G. H., Nemes, A. & Axel, R. Mice deficient in G_{olf} are anosmic. *Neuron* **20,** 69–81 (1998).
18. Zhao, H. & Reed, R. R. X inactivation of the *OCNC1* channel gene reveals a role for activity-dependent competition in the olfactory system. *Cell* **104,** 651–660 (2001).
19. Spehr, M. *et al.* Essential role of the main olfactory system in social recognition of major histocompatibility complex peptide ligands. *J. Neurosci.* **26,** 1961–1970 (2006).
20. Leinders-Zufall, T. *et al.* Contribution of the receptor guanylyl cyclase GC-D to chemosensory function in the olfactory epithelium. *Proc. Natl Acad. Sci. USA* **104,** 14507–14512 (2007).
21. Munger, S. D. *et al.* Central role of the CNGA4 channel subunit in Ca^{2+}-calmodulin-dependent odor adaptation. *Science* **294,** 2172–2175 (2001).
22. Shepherd, G. M., Chen, W. R. & Greer, C. A. in *The Synaptic Organization of the Brain* (ed. Shepherd, G. M.) 165–216 (Oxford Univ. Press, 2004).
23. Wachowiak, M. & Shipley, M. T. Coding and synaptic processing of sensory information in the glomerular layer of the olfactory bulb. *Semin. Cell Dev. Biol.* **17,** 411–423 (2006).
24. Nickell, W. T., Shipley, M. T. & Behbehani, M. M. Orthodromic synaptic activation of rat olfactory bulb mitral cells in isolated slices. *Brain Res. Bull.* **39,** 57–62 (1996).
25. Gabellec, M. M., Panzanelli, P., Sassoe-Pognetto, M. & Lledo, P. M. Synapse-specific localization of vesicular glutamate transporters in the rat olfactory bulb. *Eur. J. Neurosci.* **25,** 1373–1383 (2007).
26. Richard, M. B., Taylor, S. R. & Greer, C. A. Age-induced disruption of selective olfactory bulb synaptic circuits. *Proc. Natl Acad. Sci. USA* **107,** 15613–15618 (2010).
27. Yu, C. R. *et al.* Spontaneous neural activity is required for the establishment and maintenance of the olfactory sensory map. *Neuron* **42,** 553–566 (2004).
28. Puche, A. C. & Shipley, M. T. Odor-induced, activity-dependent transneuronal gene induction *in vitro*: mediation by NMDA receptors. *J. Neurosci.* **19,** 1359–1370 (1999).
29. Cho, J. Y., Min, N., Franzen, L. & Baker, H. Rapid down-regulation of tyrosine hydroxylase expression in the olfactory bulb of naris-occluded adult rats. *J. Comp. Neurol.* **369,** 264–276 (1996).
30. Baker, H. *et al.* Targeted deletion of a cyclic nucleotide-gated channel subunit (OCNC1): biochemical and morphological consequences in adult mice. *J. Neurosci.* **19,** 9313–9321 (1999).
31. Kobayakawa, K. *et al.* Innate versus learned odour processing in the mouse olfactory bulb. *Nature* **450,** 503–508 (2007).
32. Wesson, D. W., Levy, E., Nixon, R. A. & Wilson, D. A. Olfactory dysfunction correlates with amyloid-β burden in an Alzheimer's disease mouse model. *J. Neurosci.* **30,** 505–514 (2010).
33. Sakano, H. Neural map formation in the mouse olfactory system. *Neuron* **67,** 530–542 (2010).
34. Pyrski, M. *et al.* Sodium/calcium exchanger expression in the mouse and rat olfactory systems. *J. Comp. Neurol.* **501,** 944–958 (2007).
35. Au, W. W., Treloar, H. B. & Greer, C. A. Sublaminar organization of the mouse olfactory bulb nerve layer. *J. Comp. Neurol.* **446,** 68–80 (2002).
36. Ukhanov, K., Leinders-Zufall, T. & Zufall, F. Patch-clamp analysis of gene-targeted vomeronasal neurons expressing a defined V1r or V2r receptor: ionic mechanisms underlying persistent firing. *J. Neurophysiol.* **98,** 2357–2369 (2007).
37. Papes, F., Logan, D. W. & Stowers, L. The vomeronasal organ mediates interspecies defensive behaviors through detection of protein pheromone homologs. *Cell* **141,** 692–703 (2010).

METHODS

Human biopsies. Human nasal mucosa was obtained by biopsy during routine nasal surgery with patients under general anaesthesia. Biopsy specimens were obtained from three individuals and snap-frozen in liquid nitrogen for later processing. All samples were obtained under a protocol approved by the Ethics Committee of the University of Saarland School of Medicine. All biopsy tissues were obtained with the informed consent of the patients.

Human psychophysics. The UPSIT was obtained from Sensonics. The test was applied over a period of 25 min. Testing and scoring was done according to standardized operating procedures summarized in the test manual. The reference values have been derived from recorded reference ranges for the UPSIT test based on British individuals.

Olfactory mucosa biopsies and PCR analyses. Human surgical material containing olfactory mucosa collected from three different patients was examined individually. RT–PCRs from human samples were performed on a MyCycler (BIO-RAD) with Herculase (Agilent Technologies) following suppliers' instructions. To amplify human $G\alpha_{olf}$ we used the oligonucleotides TGGAAAGA ATCGACAGCGTCAGC and GGCCACCAACATCAAACATGTGG. Human $Na_v1.7$ was amplified by CATGAATAACCCACCGGACTG and CCTATGCC CTTCGACACCAAGG. PCR conditions were: 95 °C for 2 min pre-denaturation, followed by 35 cycles (95 °C for 30 s, 60 °C for 30 s ($G\alpha_{olf}$) or 1 min ($Na_v1.7$), 72 °C for 30 s), followed by a final extension 72 °C for 5 min. Mouse tissue was pooled from four different B6 mice (4–8 weeks old). RNA was isolated with the InnuPREP RNA isolation kit (Analyticjena). RNA quality was assessed by gel electrophoresis and photometric measurements. cDNA was synthesized from 0.5 µg of total RNA using the Smart cDNA Synthesis technology (Clontech) and Supercript II reverse transcriptase (Invitrogen). qPCR for different mouse Na_v subunits were done on a My-iQ-cycler using iQ^{TM} SYBRGreen Supermix following the supplier's instructions (BIO-RAD). We used the following oligonucleotides: $Na_v1.1$ (AGCCTGGTAGAACTTGGCCTTGC and TGCCAACCA CGGCAAAAATAAAG); $Na_v1.2$ (TGGGATCTTCACCGCAGAAATG and TGGGGCCAGGATTTTGCCAAC); $Na_v1.3$ (AGCTTGGCCTGGCAAACGTG and ATGCCGACCACGGCAAAAATG); $Na_v1.5$ (ACAGCCGAGTTTGAG GAGATGC and CGCTGATTCGGTGCCTCA); $Na_v1.6$ (ACGCCACAATTC GAACATGTCC and CCTGGCTGATCTTACAGACGCA); $Na_v1.7$ (ACGGAT GAATTCAAAAATGTACTTGCAG and GTTCTCGTTGATCTTGCAAACA CA). PCR conditions were: 95 °C for 3 min pre-denaturation, followed by 42 cycles 95 °C for 30 s, 64 °C for 20 s, 72 °C for 30 s. Each reaction was performed in three replicates on 96-well plates and analysed with the iQ5 Software (BIO-RAD). Specificity of all PCR products was confirmed by gel electrophoresis and sequencing.

Mice. Animal care and experimental procedures were performed in accordance with the guidelines established by the animal welfare committee of the University of Saarland School of Medicine. Mice were kept under a standard light/dark cycle with food and water *ad libitum*. Tissue-specific, $Na_v1.7$-deficient mice were generated by crossbreeding 'floxed' $Na_v1.7$ mice that carry two *loxP* sites, flanking exons 14 and 15 of *Scn9a*[12] with homozygous OMP–Cre mice (B6;129P2-*Omp*$^{tm4(cre)Mom}$/MomJ) that express Cre recombinase under the control of the OMP promoter[13]. Further breeding established offspring that were both homozygous for the floxed $Na_v1.7$ alleles and heterozygous for *cre* and *Omp*. In these mice, Cre-mediated $Na_v1.7$ deletion was restricted to OMP-positive cells. Additionally, C57BL/6J (B6) and OMP–GFP (B6;129P2-*Omp*^{tm3Mom}/MomJ) mice were used.

Immunohistochemistry. Perfusion of mice and preparation of mouse olfactory tissues for immunohistochemistry followed previously described methods[34]. Cryosections (10–12 µm) of either human or mouse olfactory tissues were post-fixed using 4% paraformaldehyde in PBS, before blocking and antibody administration. Primary antibodies were: mouse-specific anti-$Na_v1.7$ (1:500, rabbit polyclonal; Millipore), human-specific anti-$Na_v1.7$ (1:500, rabbit polyclonal; Abcam), $Na_v1.3$ (1:500, rabbit polyclonal; Millipore), OMP (1:3,000, goat polyclonal; gift of F. Margolis), vGluT2 (1:2,000, rabbit polyclonal; Synaptic Systems), tyrosine hydroxylase (TH, 1:3,000, mouse monoclonal; ImmunoStar). Secondary antibodies and conjugated compounds were: Alexa-Fluor 488 donkey-anti-goat (1:1,000; Invitrogen), Alexa-Fluor 555 donkey-anti-rabbit (1:1,000; Invitrogen), Alexa-Fluor 546 Streptavidin (1:200; Invitrogen). Procedures were conducted at room temperature (21 °C), except for incubation in primary antibodies (4 °C). Expression of $Na_v1.7$ in human was detected by direct immunofluorescence. Expression of $Na_v1.3$ and $Na_v1.7$ in mouse was detected by tyramid signal amplification using manufacturer's protocol (TSA-Biotin System, Perkin Elmer). Incubation in primary antibody was for 2–3 days, in biotinylated anti-rabbit antibody (1:400; Jackson ImmunoResearch) for 1 h, in streptavidin-HRP (1:100) for 30 min, in biotinylated tryamid (1:100) for 10 min, and visualized using Alexa 546-conjugated steptavidin (Invitrogen, 1:200). OMP colocalization

was detected using a Alexa 488-conjugated anti-goat secondary antibody. Detection of vGluT2 was exactly as previously described[26]. TH was detected in 30-µm free-floating sections using the avidin-biotin method (Vectastain ABC-Elite, Vector). Incubation in primary TH antibody was for 1 day, in biotinylated horse-anti-mouse secondary antibody (1:400, Vector Laboratories) for 1 h, and in avidin/biotin–HRP complex (Vector) for 90 min. Immunoreactivity was visualized with 0.05 g l^{-1} 3,3′-diaminobenzidine and 0.015% H_2O_2. Fluorescence images were acquired on either a BX71 microscope attached to a DP71 camera (Olympus) or an LSM 710/ConfoCor-3 microscope (Zeiss). Image stacks are presented as maximum intensity projections, assembled and minimally adjusted in brightness using Adobe PhotoShop 6.0.

Electron microscopy. Following routine processing for electron microscopy, as previously described[26,35], thin 70–100-nm sections were cut on a Reichert Ultracut E and examined on a JEOL 1200 transmission electron microscope. Images were captured at ×12,000, digitized at 1,200 dots per inch (DPI), and examined for ultrastructural features of the olfactory sensory axons and their synaptic terminals.

Electrophysiology. Whole-cell patch-clamp recordings from individual OSNs were obtained in acute MOE tissue slices of P1–P5 mice[19]. The anterior aspect of the head containing olfactory epithelium and bulb was embedded in agarose (4%), placed in oxygenated, ice-cold extracellular solution (95% O_2, 5% CO_2) containing: 120 mM NaCl, 25 mM NaHCO$_3$, 5 mM KCl, 5 mM BES (*N,N*-bis[2-hydroxyethyl]-2-aminoethansulphonic acid), 1 mM MgSO$_4$, 1 mM CaCl$_2$, 10 mM glucose, osmolarity adjusted to 300 mOsm, pH 7.3. Coronal slices (250 µm) were cut on a vibratome (Microm HM 650 V), transferred to a recording chamber and kept under continuous flow (2 ml min^{-1}) of oxygenated solution or remained on ice in oxygenated solution until needed (for up to 4 h). Experiments were performed at room temperature. The CsCl-based electrode solution contained: 140 mM CsCl, 1 mM EGTA, 10 mM HEPES, 0.5 mM GTP Na-salt, 2 mM ATP Mg-salt, pH 7.1, 290 mOsm. To assess OSN firing properties under non-invasive conditions, we used extracellular loose-patch recording from OSN knobs[20]. In this case, the septal epithelium of juvenile (P1–P5) or adult mice was dissected and transferred to a recording chamber. Patch pipettes (9–12 MΩ) were filled with a HEPES-based extracellular solution containing: 140 mM NaCl, 5 mM KCl, 1 mM MgCl$_2$, 1 mM CaCl$_2$, 10 mM HEPES, pH 7.4, 300 mOsm. IBMX was prepared in 10 mM stock solution containing 5% dimethylsulphoxide (DMSO) (v/ v). For M/T cell recordings, brains were rapidly dissected in ice-cold oxygenated (95% O_2, 5% CO_2) solution containing: 83 mM NaCl, 26.2 mM NaHCO$_3$, 1 mM NaH$_2$PO$_4$, 2.5 mM KCl, 3.3 mM MgSO$_4$, 0.5 mM CaCl$_2$, 70 mM sucrose, pH 7.3, 300 mOsm. Horizontal olfactory bulb slices (300 µm) were cut in this solution. Until use, slices were transferred to oxygenated modified artificial cerebrospinal fluid (ACSF, 95% O_2, 5% CO_2) containing: 125 mM NaCl, 25 mM NaHCO$_3$, 2.5 mM KCl, 1.25 mM NaH$_2$PO$_4$, 1 mM MgCl$_2$, 2 mM CaCl$_2$ and 25 mM glucose. Recording pipettes had resistances of 4–7 MΩ. M/T cells were identified by size and location of their somata and filled with Lucifer Yellow during patch recording. The intracellular solution contained: 140 mM KCl, 1 mM EGTA, 10 mM HEPES, 1 mM ATP Na-salt, 0.5 mM GTP Mg-salt, 0.1 mM Lucifer Yellow; pH 7.1, 290 mOsm. M/T cells were held at −55 to −60 mV. Input and series resistances were 200–300 MΩ and 15–20 MΩ, respectively. After establishing a whole-cell recording, the ONL was stimulated using a glass electrode (1–1.5 MΩ) filled with HEPES-buffered extracellular solution connected to an electrical stimulator (single stimulus: 20 ms, 40 V, 266–400 µA). The stimulus pipette was placed rostrally to the recorded cell in the ONL. If a given M/T cell showed no postsynaptic response, the position of the stimulus pipette was changed until OSN axon bundles were found that caused M/T cell responses. Ionic currents were analysed using PulseFit 8.54 (HEKA) and IGOR Pro software (Wavemetrics)[36]. OSNs with leak currents >20 pA and M/T cells with leak currents >100 pA (all measured at −70 mV) were excluded from analysis. Cell capacitance (C_m) was monitored using the automated function of the EPC-9 amplifier. A stable C_m value over time was an important criterion for the quality of an experiment. Spike analysis was done off-line using IGOR Pro software with custom-written macros. Chemicals were purchased from Sigma unless otherwise stated. Drugs used in the electrophysiological experiments were prepared as stock solutions in DMSO or distilled water and diluted to the final concentration in HEPES-based extracellular solution. NaCl, MgCl$_2$, glucose and CaCl$_2$ were from Merck. IBMX (100 µM) and cineole (100 µM) were diluted in a HEPES-buffered extracellular solution (<0.1% DMSO) and focally ejected using multibarrel stimulation pipettes.

Behavioural tests. The innate olfactory preference test followed previously described procedures[31]. Briefly, mice were habituated to the test conditions before odour exposure. Mice were individually placed in an empty cage for 30 min and then transferred to a new cage. This habituation was repeated three to four times for each animal. Soon after habituation, mice were transferred to the test cage, and a filter paper scented with a test odorant was introduced. Investigation times of the filter paper during the 3-min test period was recorded and quantified. Odour

stimuli were freshly collected male and female B6 mouse urine (5 μl), peanut butter (10% w/v, 15 μl), milk powder (10% w/v, 15 μl), water (15 μl) and cineole (100 μM, 15 μl).

For the innate olfactory avoidance test, following habituation (see innate preference test), a filter paper scented with 5 μl TMT (7.6 mM) was placed in one corner of the test cage. Mouse behaviour was recorded for 30 min. The test cage was subdivided into three equally sized areas. Time spent in area 1 of the cage (farthest distance from the TMT source) was evaluated as avoidance, whereas time spent in area 2 (consisting of the TMT source) was evaluated as attraction[37]. Animal movements were tracked with SwisTrack (Swarm Intelligent Systems Group, Swiss Federal Institute of Technology).

For the olfactory habituation–dishabituation assay, following habituation (see innate preference test) mice were exposed for 3 min to distilled water (15 μl). This procedure was repeated three times with 1-min intervals, followed by a three-time presentation of female urine (5 μl) and a three-time presentation of male urine (5 μl). Investigation times during the 3-min test periods were measured.

For the pup retrieval test, lactating mice were habituated to the experiment for several minutes. Experiments were performed in the bedded home cages of the dams. Three pups (1–3-days old) were removed from the nest and randomly distributed in the cage. The latency for pup retrieval back into the nest was measured. If a dam had not completed retrieval within 10 min the test was terminated, resulting in a latency of 600 s.

Experiments were performed in empty standard cages ($38 \times 19 \times 12$ cm) and test substances were applied on filter paper ($\sim 1 \times 3$ cm). Mouse behaviour was recorded with a digital camera (Sony) for the experimental times indicated. Statistical video analyses were done randomly and blindly. Peanut butter (Barney's Best) and milk powder (Bio-Anfangsmilch, Hipp) were diluted to 10% (w/v) in water.

IGF-1 assays. IGF-1 levels were measured by sandwich ELISA (ALPCO Diagnostics). IGF-1 was dissociated from the binding proteins by diluting samples with an acidic buffer. The analytical sensitivity of the assay was 0.029 ng ml^{-1}. Inter and intra-assay variability was below 7%. Experiments used plasma of 4–5-weeks-old mice ($n = 4$, each genotype).

Statistics. Data were analysed using NCSS 2004 statistical software (NCSS). The Student's t-test (two-tailed) was used for measuring the significance of difference between two distributions. Multiple groups were compared using a one-way or two-way analysis of variance (ANOVA) with Fisher's LSD as a post hoc comparison. Unless otherwise stated, results are presented as means ± s.e.m.

ARTICLE

Cortical representations of olfactory input by trans-synaptic tracing

Kazunari Miyamichi[1], Fernando Amat[2], Farshid Moussavi[2], Chen Wang[3], Ian Wickersham[4]†, Nicholas R. Wall[4], Hiroki Taniguchi[5], Bosiljka Tasic[1], Z. Josh Huang[5], Zhigang He[3], Edward M. Callaway[4], Mark A. Horowitz[2] & Liqun Luo[1]

In the mouse, each class of olfactory receptor neurons expressing a given odorant receptor has convergent axonal projections to two specific glomeruli in the olfactory bulb, thereby creating an odour map. However, it is unclear how this map is represented in the olfactory cortex. Here we combine rabies-virus-dependent retrograde mono-trans-synaptic labelling with genetics to control the location, number and type of 'starter' cortical neurons, from which we trace their presynaptic neurons. We find that individual cortical neurons receive input from multiple mitral cells representing broadly distributed glomeruli. Different cortical areas represent the olfactory bulb input differently. For example, the cortical amygdala preferentially receives dorsal olfactory bulb input, whereas the piriform cortex samples the whole olfactory bulb without obvious bias. These differences probably reflect different functions of these cortical areas in mediating innate odour preference or associative memory. The trans-synaptic labelling method described here should be widely applicable to mapping connections throughout the mouse nervous system.

The functions of mammalian brains are based on the activity patterns of large numbers of interconnected neurons that form information processing circuits. Neural circuits consist of local connections—where pre- and postsynaptic partners reside within the same brain area—and long-distance connections, which link different areas. Local connections can be predicted by axon and dendrite reconstructions[1], and confirmed by physiological recording and stimulation methods[2]. Long-distance connections are more difficult to map, as commonly used methods can only trace bulk projections with a coarse resolution. Most methods cannot distinguish axons in passing from those that form synapses, or pinpoint the neuronal types to which connections are made[1,2]. Trans-synaptic tracers can potentially overcome these limitations[2]. Here we combine a retrograde rabies-virus-dependent mono-trans-synaptic labelling technique[3] with genetic control of the location, number and cell type of 'starter' neurons to trace their presynaptic partners. We systematically mapped long-distance connections between the first olfactory processing centre, the olfactory bulb, and its postsynaptic targets in the olfactory cortex including the anterior olfactory nucleus (AON), piriform cortex and amygdala (Supplementary Fig. 1).

Genetic control of trans-synaptic tracing

Rabies virus can cross synapses from postsynaptic to presynaptic neurons with high specificity[4], without notable defects in the morphology or physiology of infected neurons for extended periods of time[3,5]. Recent genetic modifications of rabies virus have permitted mono-trans-synaptic labelling[3]. Specifically, the rabies envelope glycoprotein (G) required for viral spread was replaced with a fluorescent marker[6]. Further, the virus was pseudotyped with EnvA, an avian virus envelope protein that lacks an endogenous receptor in mammals, and thus cannot infect wild-type mammalian cells. However, it can infect cells expressing the EnvA receptor TVA, and can subsequently produce infectious particles if TVA-expressing cells also

express G to complement the ΔG rabies virus (Fig. 1a, bottom). The new viral particles can cross synapses to label presynaptic partners of starter neurons. As trans-synaptically infected neurons do not express G, the modified virus cannot spread from them to other neurons. Paired recordings in cultured brain slices support the efficacy and specificity of this strategy[3].

To extend this method to a limited number of starter cells of a defined type and at a precise location in vivo, we combined mouse genetics and viral infections (Fig. 1a, b). We created a transgenic mouse (CAG-stop-tTA2) that conditionally expresses the tetracycline transactivator tTA2 under the control of a ubiquitous CAG promoter only upon Cre-mediated excision of a transcriptional stop cassette. After crossing these mice with transgenic mice expressing the tamoxifen-inducible Cre (CreER), a small fraction of CreER+ cells also express tTA2 following tamoxifen induction. We then used stereotactic injections to deliver into specific regions of the brain an adeno-associated virus (AAV) serotype 2 expressing three proteins: histone–GFP, TVA and G, under the control of a tetracycline-response element (TRE). Expression of TVA and G allows infected, tTA2+ cells to be receptive to infection by the modified rabies virus, which we injected into the same location two weeks later. We define starter cells as those infected by both AAV and rabies virus, and therefore labelled by both histone–GFP and mCherry; their presynaptic partners are infected only by rabies virus and therefore express only mCherry.

We tested our strategy by using a ubiquitously expressing actin-CreER[7] in combination with CAG-stop-tTA2 in the neocortex. Starter cells could be unambiguously identified by histone–GFP expression (Supplementary Fig. 2). In all but one case, we observed more than one starter cell (Supplementary Fig. 3 shows the example of a single starter cell). In a typical example, 35 starter cells in the motor cortex expressed histone–GFP and mCherry (Fig. 1c (3)), demonstrating that AAV and rabies virus can infect the same cells in vivo. In addition to many locally labelled cells (Fig. 1c (1)), mCherry+ cells were

[1]HHMI/Department of Biology, Stanford University, Stanford, California 94305, USA. [2]Department of Electrical Engineering, Stanford University, Stanford, California 94305, USA. [3]Department of Neuroscience, Children's Hospital, Boston, Massachusetts 02115, USA. [4]Systems Neurobiology Laboratory, The Salk Institute for Biological Studies and Neurosciences Graduate Program, University of California, San Diego, La Jolla, California 92037, USA. [5]Cold Spring Harbor Laboratory, New York, New York 11724, USA. †Present address: HHMI/Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
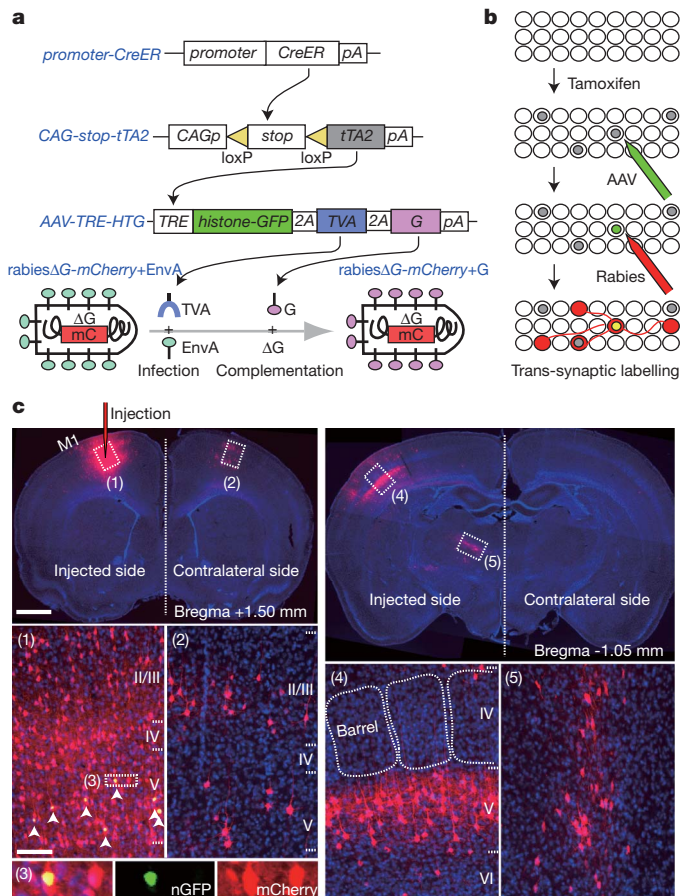
Figure 1 | Genetic control of rabies-mediated neural circuit tracing.
a, **b**, Schematic representation of the methodology used to control the location, number and type of starter cells for rabies-virus-mediated trans-synaptic labelling. tTA2 is expressed in a small subset of CreER[+] cells (grey nuclei in **b**). tTA2 activates an AAV-delivered transgene to express: (1) a histone–GFP marker to label the nuclei of starter cells in green; (2) EnvA receptor (TVA) to enable subsequent infection by EnvA-pseudotyped rabies virus (rabiesΔG-mCherry+EnvA); and (3) rabies glycoprotein (G) to initiate trans-synaptic labelling. **c**, Top left, a 60-μm coronal section that includes the injection site in the motor cortex (M1). Cells labelled with both histone–GFP (nGFP) and mCherry (arrowheads in **c** (1), magnified in **c** (3)) can be distinguished from cells labelled with mCherry alone, which are found near the injection site (**c** (1)), in the contralateral motor cortex (**c** (2)), in the somatosensory barrel cortex (top right; magnified in **c** (4)), and in the motor-specific ventrolateral nucleus of the thalamus (**c** (5)). Scale bars, 1 mm for low-magnification images at the top, 100 μm for high-magnification images at the bottom.

enriched in layers II, III and V in the contralateral motor cortex (Fig. 1c (2)), consistent with layer specificity of callosal projections[8]. mCherry[+] cells were also found in layers III and V of the ipsilateral somatosensory cortex (Fig. 1c (4)) and in motor-specific thalamic nuclei (Fig. 1c (5)), which are known sources of monosynaptic inputs to the motor cortex[9].

In all experiments, histone–GFP[+] cells were found within 450 μm of the injection sites, consistent with a previous report that AAV serotype 2 predominantly infects neurons locally[10]. Omitting AAV or tamoxifen yielded no trans-synaptically labelled neurons (Supplementary Fig. 4). Moreover, our strategy labelled neurons only through synaptic connections but not through axons in passage (Supplementary Fig. 5). Finally, rabies virus spread was restricted to neurons directly connected to starter cells, and only in the retrograde direction (Supplementary Fig. 6). Together, these experiments validated our genetic strategy for retrograde mono-trans-synaptic labelling *in vivo*.

AON maintains the dorsal–ventral topography

In the mouse, olfactory receptor neurons that express a single type of odorant receptor send convergent axonal projections to a specific pair of glomeruli in the lateral and medial olfactory bulb[11–13]. Odorants are detected by combinations of olfactory receptor neuron classes[14], and are represented as spatiotemporal activity patterns of glomeruli[15]. Each mitral cell sends its apical dendrite to a single glomerulus and thus receives direct input from a single olfactory receptor neuron class. Mitral cell axons relay information to the olfactory cortex (Supplementary Fig. 1a). Previous axon tracing studies showed that individual mitral cells send axons to distinct cortical areas, and that small cortical regions receive broad input from the olfactory bulb[16–19]. However, understanding the principles underlying odour perception and odour-mediated behaviours requires systematic and quantitative analysis of connection patterns of mitral cells with their cortical target neurons.

We first established that mitral cells throughout the olfactory bulb can be infected by rabies virus via their axons (Supplementary Fig. 7). We then applied our strategy (Fig. 1a, b) to specific areas of the AON, piriform cortex and cortical amygdala (Supplementary Fig. 1b), and examined the distribution of trans-synaptically labelled mitral cells. In a typical example, 11 clustered starter cells in the AON (Fig. 2a) resulted in 69 labelled mitral cells distributed widely across the olfactory bulb (Supplementary Fig. 8 and Supplementary Movie 1). Bright mCherry fluorescence from rabies virus allowed us to unequivocally follow the primary dendrites of the labelled mitral cells to single target glomeruli (Fig. 2b). Each mitral cell sent its apical dendrite into a single glomerulus. Four glomeruli were each innervated by two labelled mitral cells (Fig. 2b, right, and Supplementary Table 1).

To quantitatively compare the patterns of labelled glomeruli from different animals, we established a three-dimensional (3D) reconstruction protocol for the olfactory bulb, and aligned each olfactory bulb to a standard olfactory bulb model (Fig. 2c). To test the accuracy of this procedure, we reconstructed and aligned olfactory bulbs from three *P2-IRES-tauGFP* transgenic mice[20]. These GFP-labelled glomeruli were located within a distance of a few glomeruli from each other (Supplementary Fig. 9), consistent with the natural variability of olfactory receptor neuron axon targeting[21]. This precision of our 3D reconstruction enables the comparison of olfactory bulbs from different animals.

The AON has been proposed to provide feedforward modification of information from the olfactory bulb to the piriform cortex[22]. Little is known about its organization except for a small and distinct AON pars externa, which maintains dorsal–ventral olfactory bulb topography[23–25]. We injected AAV and rabies virus to different areas of the AON (Supplementary Table 1), and established an AON 3D-reconstruction protocol analogous to that for the olfactory bulb (Fig. 2c, left). Labelled glomeruli from AON injections were distributed widely in the olfactory bulb (Fig. 2c, middle). However, starter cells from the ventral and dorsal AON preferentially labelled ventral and dorsal glomeruli, respectively (Fig. 2c). To quantify the spatial distributions of starter cells in the AON and trans-synaptically labelled glomeruli in the olfactory bulb, we introduced a cylindrical coordinate system into the olfactory bulb and AON models, where Z represents the position along the anterior–posterior axis and θ represents the angle from the polar axis (Fig. 2c). No correlations were found between Z_{AON} and θ_{OB} (where OB is olfactory bulb), Z_{AON} and Z_{OB}, or θ_{AON} and Z_{OB} (Supplementary Fig. 10a). However, we found a strong positive correlation ($R^2 = 0.79$) between θ_{AON} and θ_{OB} (Fig. 2d), which correspond to the dorsal–ventral axes of the AON and olfactory bulb, respectively. Thus, the AON maintains the dorsal–ventral topography of the olfactory bulb.

A coarse topography exists between olfactory receptor neuron cell-body positions in the olfactory epithelium and target glomeruli in the olfactory bulb along the dorsal–ventral axis[26]. Specifically, the olfactory cell adhesion molecule (OCAM) is expressed in a subset of olfactory receptor neurons[27] that project to the ventral ~55% of glomeruli in the olfactory bulb. In the olfactory bulb, ~25° clockwise rotation of the
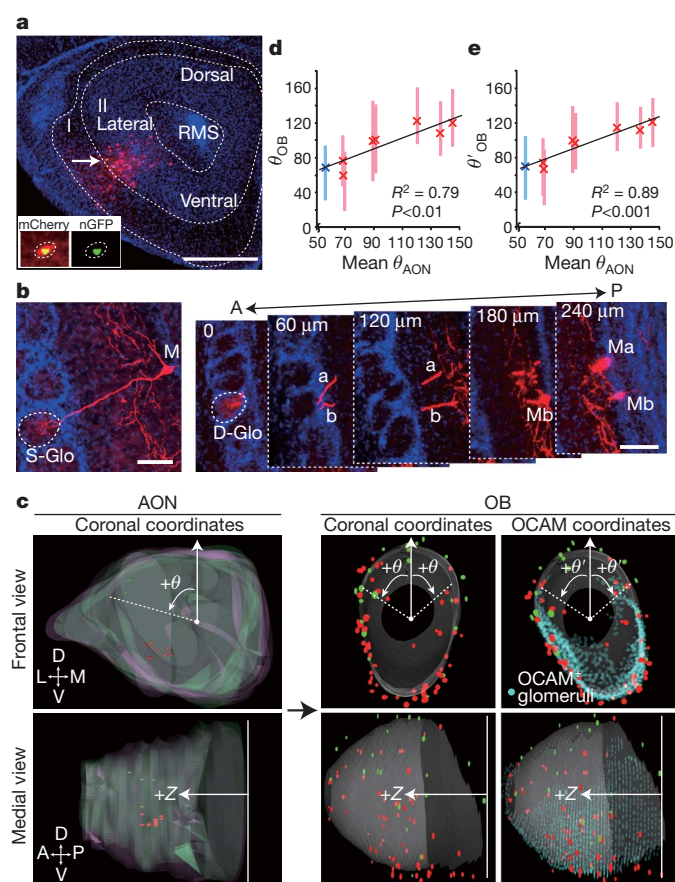
Figure 2 | The olfactory bulb to AON connections show a dorsal–ventral topography. a, A 60-μm coronal section with two starter cells located in layer II of the ventrolateral AON, one of which (arrow) is magnified in the inset. RMS, rostral migratory stream. Scale bar, 500 μm. **b,** Typical examples of trans-synaptically labelled mitral cells from cortical starter cells. Left, a 60-μm coronal section that captures both the cell body and the apical dendrite of a mitral cell. Right, more frequently, a mitral cell apical dendrite spans several consecutive 60-μm coronal sections. S-Glo, glomerulus innervated by a single labelled mitral cell (M). D-Glo, glomerulus innervated by two labelled mitral cells (Ma, Mb). A, anterior; P, posterior. Scale bar, 100 μm. **c,** Superimposed 3D reconstructions of the AONs and olfactory bulbs (OBs) from two injected brains. Eleven red and four green starter cells from two AONs labelled red and green glomeruli, respectively. Light red and green, contours of two superimposed AONs. D, dorsal; L, lateral; M, medial; V, ventral. **d, e,** Correlations between θ_{AON} and θ_{OB} (**d**) and θ_{AON} and θ'_{OB} (**e**). Crosses represent mean θ_{AON} (x-axis) and mean θ_{OB} or θ'_{OB} (y-axis). Error bars represent 50% of the distribution surrounding the mean θ_{OB} or θ'_{OB}. R^2, square of Pearson's correlation coefficient; P, statistical significance tested against the null hypothesis assuming no correlation between θ_{AON} and θ_{OB} or θ'_{OB}. Red and blue, experiments using *actin-CreER* and *CaMKII-CreER*[47], respectively.

polar axis around the z-axis maximized the separation of OCAM$^+$ and OCAM$^-$ glomeruli (Supplementary Fig. 11). In this new OCAM coordinate system represented by θ'_{OB} (Fig. 2c, right), the correlation coefficient between θ'_{OB} and θ_{AON} increased to $R^2 = 0.89$ (Fig. 2e), showing that adjusting the dorsal–ventral axis of the olfactory bulb according to a biological marker improved the AON and olfactory bulb topographic correspondence. Thus, the topography between the olfactory epithelium and the olfactory bulb further extends to the AON.

Dorsally biased olfactory bulb input to amygdala

Mitral cell axons project to the anterior and posterolateral cortical amygdala[28,29]. The organization of this axonal input is unknown. We injected AAV and rabies virus to small areas within these regions, and mapped starter cells onto a common schematic drawing based on

anatomical landmarks (Fig. 3a). Trans-synaptically labelled mitral cells and glomeruli from amygdala starter cells were broadly distributed in the olfactory bulb. However, the labelled glomeruli were enriched in the dorsal olfactory bulb (Fig. 3c). For quantification, we compared the mean experimental θ'_{OB} for each injection with mean θ'_{OB} values produced by computer simulation from the same number of glomeruli distributed randomly throughout the olfactory bulb ($^{sim}\theta'_{OB}$). For the AON experiments, the mean experimental θ'_{OB} values for the majority of the samples were significantly larger or smaller than the corresponding mean $^{sim}\theta'_{OB}$ values (Fig. 3e, left), reflecting the dorsal–ventral topography between the olfactory bulb and the AON. By contrast, none of the mean θ'_{OB} values from the amygdala was significantly larger than the corresponding mean $^{sim}\theta'_{OB}$ (Fig. 3e, middle). Six out of ten mean θ'_{OB} values from the cortical amygdala fell significantly below the corresponding mean $^{sim}\theta'_{OB}$ values. For these dorsally biased samples, the density of labelled glomeruli gradually decreased along the dorsal–ventral axis without a sharp boundary (Supplementary Fig. 12). Simple spatial correspondence between starter-cell locations and the degree of dorsal bias was not evident (Supplementary Fig. 10b). In summary, the cortical amygdala overall receives biased input from the dorsal olfactory bulb.

Less organized olfactory bulb input to piriform cortex

The piriform cortex is the largest cortical area in the olfactory cortex. Recent physiological analysis[30,31] found that neurons activated by specific odours are apparently not spatially organized; the underlying anatomical basis is unclear. We injected AAV and rabies virus into several areas in the anterior and posterior piriform cortex, and mapped starter cells from different brains onto a common schematic drawing of the entire piriform cortex based on anatomical landmarks (Fig. 3b). Labelled glomeruli were broadly distributed throughout the olfactory bulb, regardless of starter-cell locations in the piriform cortex (Fig. 3d). In sharp contrast to trans-synaptic labelling from the AON or amygdala, where different samples showed highly variable mean θ'_{OB}, mean θ'_{OB} values from the piriform cortex tracings were much less variable and closely resembled a random distribution (Fig. 3e). Only one out of ten samples had a mean θ'_{OB} slightly above the 95th percentile of the mean $^{sim}\theta'_{OB}$. Further, no strong spatial correspondence was evident in correlation analyses of θ'_{OB}, Z_{OB} and the location of starter cells in the piriform cortex (Supplementary Fig. 10c). These data indicate that highly restricted areas of the piriform cortex receive direct mitral cell input representing glomeruli that are distributed throughout the olfactory bulb with no apparent spatial organization.

Convergence of mitral cell input

Convergent inputs from different glomeruli to individual cortical neurons could allow the olfactory cortex to integrate combinatorial odour representations in the olfactory bulb. In support of this, previous studies have shown that odour receptive ranges of AON cells are broader than those of mitral cells[32], and that some piriform cortex neurons are activated by a binary odour mix but not individual components[31]. However, a large fraction of inputs in these studies could come from other cortical neurons through extensive recurrent connections (Figs 2a and 3a, b). Direct convergence of mitral cell axons onto individual cortical neurons is implied in physiological studies of piriform cortical neurons in slices[33–35]. Our trans-synaptic labelling enabled a direct examination of mitral cell convergence to individual cortical neurons *in vivo*.

The convergence index, defined by the number of labelled mitral cells divided by the number of the starter cells in the cortex, exceeded 1 in all experiments using *actin-CreER* (Fig. 4a and Supplementary Table 1). This finding demonstrates that individual cortical neurons receive direct inputs from multiple mitral cells *in vivo*. As the vast majority of labelled mitral cells corresponded to different glomeruli (Supplementary Table 1), individual cortical neurons must receive direct inputs representing multiple glomeruli. This convergence index
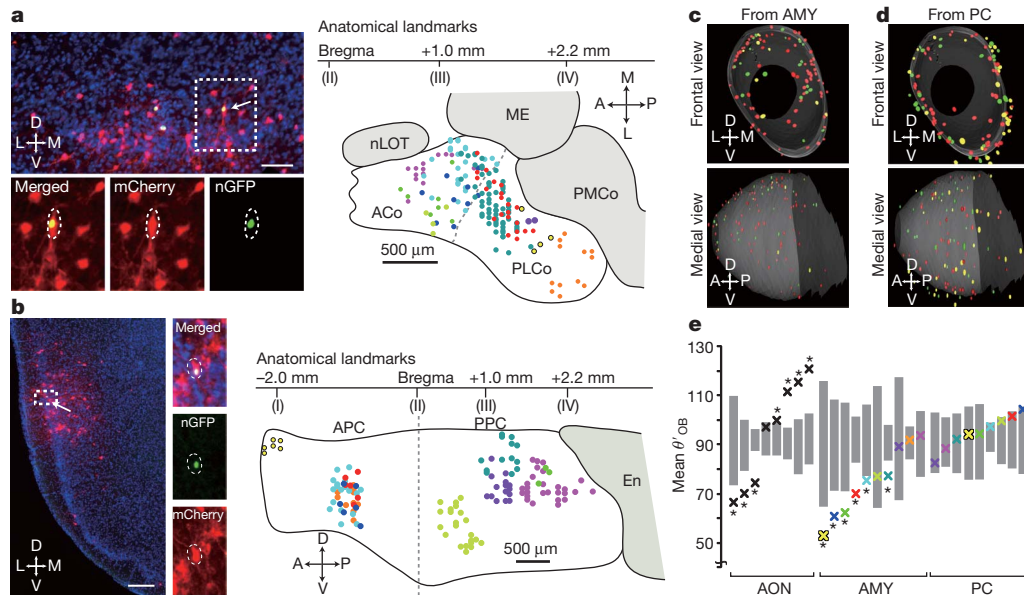
Figure 3 | Representations of olfactory bulb input in the amygdala and piriform cortex. a, b, Starter cells from the cortical amygdala and piriform cortex. Left, single coronal sections at the injection sites in the posterolateral cortical amygdala (**a**) and the posterior piriform cortex (**b**). Arrows point to starter cells magnified in insets. Scale bars, 100 μm in **a**, 200 μm in **b**. Right, schematic representations of ten independent injections each into amygdala (**a**) or piriform cortex (**b**). Starter cells from each injection are labelled with a specific colour. The dotted line denotes the rough border between the anterior cortical amygdala (ACo) and the posterolateral cortical amygdala (PLCo) based on anatomical landmarks according to a mouse brain atlas[48]. APC, anterior piriform cortex; En, lateral entorhinal cortex; ME, medial amygdala; nLOT,

nucleus of lateral olfactory tract; PMCo, posteromedial cortical amygdala; PPC, posterior piriform cortex. **c, d,** Superposition of three independent 3D reconstructions of glomerular maps with starter cells from the cortical amygdala (AMY; **c**) or the piriform cortex (PC; **d**). **e,** Mean θ'_{OB} values (crosses) from each experiment are plotted in the same column with the 95% confidence intervals for corresponding $^{sim}\theta'_{OB}$ values (grey bars). Samples with experimental mean θ'_{OB} outside the 95% confidence intervals are labelled with asterisks (*$P < 0.05$). Colours in **a** (scheme), **c** and **e** (amygdala) are matched to represent the same samples, and so are the colours in **b** (scheme), **d** and **e** (piriform cortex).

is probably an underestimate, as not all starter cells necessarily received direct mitral cell input (overestimation of the denominator), and not all cells presynaptic to starter cells were trans-synaptically infected by the rabies virus (underestimation of the numerator; see Supplementary Fig. 3).

The convergence indices varied widely in different experiments, and did not differ substantially in the three cortical areas we examined. However, in experiments that contained starter cells located in layer I, which is mostly composed of GABAergic local interneurons[36], the convergence indices were greater (Fig. 4a, red). Assuming all starter cells in a given layer contribute equally to mitral cell labelling, multiple regression analyses indicate that layer I neurons receive direct input from more mitral cells than layer II/III neurons (Fig. 4b).

To confirm the higher convergence index for layer I GABAergic neurons, we replaced the ubiquitous *actin-CreER* with *GAD2-CreER*, which is expressed only in GABAergic interneurons (Supplementary Fig. 13). We found that GABAergic neurons located in layer II or III of the piriform cortex received little direct mitral cell input, whereas those located in layer I showed a much greater convergence index (Fig. 4b, right, and Supplementary Table 1). Thus, cortical GABAergic neurons are highly diverse with respect to mitral cell innervation. These observations are in accordance with recent physiological studies[30,34], and suggest different physiological roles for these GABAergic neurons; layer I and deeper layer GABAergic neurons provide global feedforward and feedback inhibition to cortical pyramidal neurons, respectively.

Sister mitral cells connect independently

Each glomerulus is innervated by the apical dendrites of ~25 electrically coupled mitral cells[37]. We refer to these cells as 'sister' mitral cells. Sister mitral cells may preferentially connect to the same cortical postsynaptic target neurons compared to 'non-sister' mitral cells that receive direct input from different glomeruli. Such an organization could increase the signal-to-noise ratio in information transmission

from mitral cells to cortical neurons. Alternatively, sister mitral cells may connect to cortical neurons independently to deliver olfactory information widely across different cortical neurons.

We used the frequency of dually labelled glomeruli from our data set and statistical simulation to distinguish between these possibilities. Dually labelled glomeruli (D) could result from a single starter cell (Ds) or two starter cells (Dt). Assuming that an individual starter cell can receive input from any of the 2,000 glomeruli, we compared the distribution of Ds derived from our data and from a simulation according to the null hypothesis that sister mitral cells connect independently with postsynaptic targets. If sister mitral cells share significantly more postsynaptic targets than at random, then the 'data Ds' distribution should be significantly higher than the simulated 'random Ds' distribution. In all but two cases, these two distributions were not statistically different (Fig. 4c). Both exceptions came from trans-synaptic labelling from the AON, which showed dorsal–ventral topography, so the original assumptions were not accurate. When we reduced the number of accessible glomeruli to 1,500, no sample showed significant differences. Thus, our analysis indicates that individual mitral cells innervating the same glomerulus act independently in making connections with their cortical targets.

Discussion

Our study revealed several general principles that define cortical representations of the olfactory bulb input. First, individual cortical neurons receive direct input from mitral cells originating from multiple glomeruli. On average, each excitatory neuron receives direct input from four mitral cells, but this number is likely to be an underestimate. Convergence of mitral cell inputs enables cortical neurons to integrate information from discrete olfactory channels. The lower bound of four already affords ~10^{12} glomerular combinations for 1,000 olfactory channels, far exceeding the number of neurons in the mouse olfactory cortex. Thus, the olfactory cortical neuron repertoire samples only a
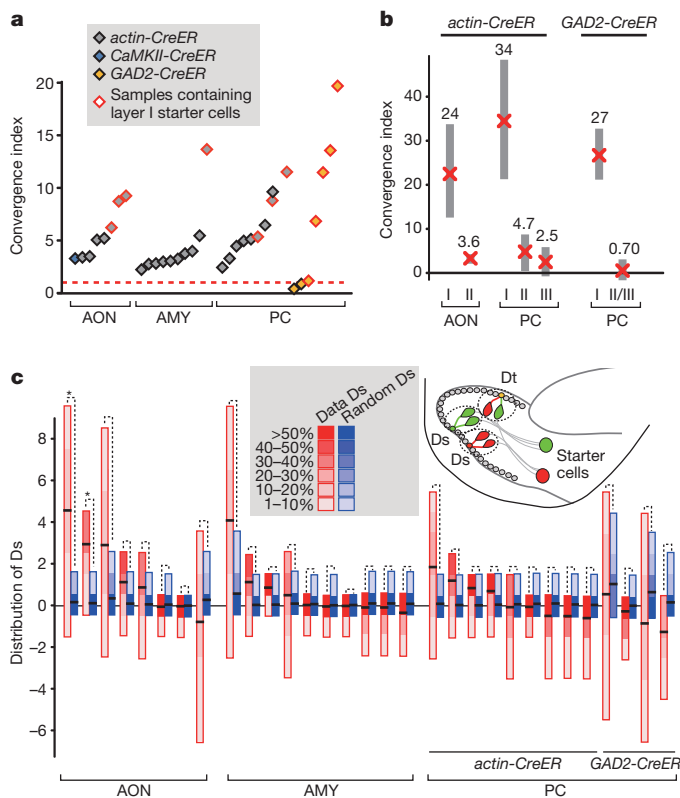
Figure 4 | Convergence and independence of mitral cell inputs.
a, Convergence index for each cortical injection experiment is represented by a diamond, with the type and layer of starter cells specified by the colour code above. AMY, amygdala; PC, piriform cortex. **b**, Multiple regression analysis to estimate the convergence indices of starter cells located in different layers of the AON and piriform cortex. Estimated mean convergence indices (red crosses) and the corresponding 95% confidence intervals (grey bars) are shown. Data from *actin-CreER* and *GAD2-CreER* were analysed separately. Injections into amygdala produced only one sample that contained layer I cells and were therefore excluded. **c**, Schematic of dually labelled glomeruli (D) resulting from two starter cells (Dt) or a single starter cell (Ds). Comparison of the distributions of Ds derived from experimentally observed frequency of D (Data Ds; red) and from simulated D based on the null hypothesis detailed in Methods (Random Ds; blue). For each sample, the distributions of 'data Ds' and 'random Ds' are shown by coloured heat maps. *$P < 0.05$.

small fraction of all possible combinations of direct olfactory bulb inputs.

Second, neurons restricted to small olfactory cortical regions receive input from glomeruli that are broadly distributed in the olfactory bulb. Although similar findings were reported previously[16–19], our study provides a higher resolution analysis of direct connectivity between mitral cells and cortical neurons, rather than inferring connection from the presence of axons, which could be a major caveat of previous tracing studies (see Supplementary Fig. 5). At the same time, mitral cells representing the same glomerulus connect independently to post-synaptic cortical neurons, thus maximizing the spread of olfactory information originating from individual olfactory channels. Our finding is consistent with analyses of axon arborization patterns of singly labelled mitral cells (S. Ghosh and colleagues; manuscript submitted).

Third, different cortical areas receive differentially organized olfactory bulb input (Supplementary Fig. 1c). The AON maintains a coarse topography along the dorsal–ventral axis, suggesting a pre-processing role for olfactory-bulb-derived information before sending to other cortical areas. A lack of apparent spatial organization in the piriform cortex with regard to olfactory bulb input provides an anatomical basis for recent physiological studies[30,31], and suggests that the piriform cortex acts as an association cortex[31,38]. In the cortical amygdala, many neurons seem to receive strongly biased input from the dorsal olfactory bulb. Mice lacking

olfactory receptor neurons that project to the dorsal olfactory bulb lose their innate avoidance of odours from predator urine and spoiled food, despite retaining the ability to sense these odours[39]. The cortical amygdala may preferentially process the olfactory information that directs innate behaviours. Our study is in agreement with similar findings using axon tracings from individual glomeruli (D. L. Sosulski and colleagues; manuscript submitted).

Interestingly, axonal arborization patterns of *Drosophila* olfactory projection neurons (equivalent to mitral cells) in higher olfactory centres show a similar organizational principle. Projection neuron axon arborization patterns in the lateral horn—a processing centre directing odour-mediated innate behaviour—are highly stereotyped with respect to projection neuron classes[40,41], and are partitioned according to the biological significance of the odorants[42]. Arborization patterns of axon collaterals of the same projection neurons in the mushroom body, an olfactory memory centre[43], are much less stereotyped[40,42], consistent with a physiological study indicating non-stereotyped connections[44]. Therefore, from insects to mammals, a common theme emerges for the representations of olfactory information: more stereotyped and selective representation of odours is necessary for directing innate behaviours, whereas broader and less stereotyped sampling of the whole olfactory space is better suited for brain regions implicated in associative memories.

The genetically controlled mono-trans-synaptic tracing described here should be widely applicable for mapping neuronal circuitry throughout the mouse brain. It is currently unknown how rabies virus crosses synapses, and whether the efficiency and specificity vary with cell type, connection strength and activity[3,5,45]. Further applications of these trans-synaptic methods to other neurons and circuits[46] will be necessary to address these questions. Nevertheless, the control experiments (Fig. 1 and Supplementary Figs 2–6) confirmed that our strategy labels neurons that are directly presynaptic to starter cells but not neurons whose axons pass through the injection sites without making synapses. Our method will be especially valuable for analysing long-distance connections that are usually refractory to physiological mapping strategies[2]. This method can be further extended to genetic manipulation of starter cells to combine circuit tracing with genetic loss- or gain-of-function experiments. These approaches will facilitate the investigation of not only the organization of information flow within neural circuits, but also the molecular basis of neuronal connections at single-cell resolution *in vivo*.

METHODS SUMMARY

Detailed methods on the generation of *CAG-stop-tTA2* mice, viral preparations, animal surgery, tissue processing, 3D reconstruction and quantitative analyses can be found in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Cowan, W. M. The emergence of modern neuroanatomy and developmental neurobiology. *Neuron* **20,** 413–426 (1998).
2. Luo, L., Callaway, E. M. & Svoboda, K. Genetic dissection of neural circuits. *Neuron* **57,** 634–660 (2008).
3. Wickersham, I. R. *et al.* Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* **53,** 639–647 (2007).
4. Ugolini, G. Specificity of rabies virus as a transneuronal tracer of motor networks: transfer from hypoglossal motoneurons to connected second-order and higher order central nervous system cell groups. *J. Comp. Neurol.* **356,** 457–480 (1995).
5. Ugolini, G. Advances in viral transneuronal tracing. *J. Neurosci. Methods* **194,** 2–20 (2010).
6. Wickersham, I. R., Finke, S., Conzelmann, K. K. & Callaway, E. M. Retrograde neuronal tracing with a deletion-mutant rabies virus. *Nature Methods* **4,** 47–49 (2007).
7. Guo, C., Yang, W. & Lobe, C. G. A Cre recombinase transgene with mosaic, widespread tamoxifen-inducible action. *Genesis* **32,** 8–18 (2002).
8. Wise, S. P. & Jones, E. G. The organization and postnatal development of the commissural projection of the rat somatic sensory cortex. *J. Comp. Neurol.* **168,** 313–343 (1976).

9. Iriki, A., Pavlides, C., Keller, A. & Asanuma, H. Long-term potentiation in the motor cortex. *Science* **245**, 1385–1387 (1989).
10. Taymans, J. M. *et al.* Comparative analysis of adeno-associated viral vector serotypes 1, 2, 5, 7, and 8 in mouse brain. *Hum. Gene Ther.* **18**, 195–206 (2007).
11. Vassar, R. *et al.* Topographic organization of sensory projections to the olfactory bulb. *Cell* **79**, 981–991 (1994).
12. Ressler, K. J., Sullivan, S. L. & Buck, L. B. Information coding in the olfactory system: evidence for a stereotyped and highly organized epitope map in the olfactory bulb. *Cell* **79**, 1245–1255 (1994).
13. Mombaerts, P. *et al.* Visualizing an olfactory sensory map. *Cell* **87**, 675–686 (1996).
14. Malnic, B., Hirono, J., Sato, T. & Buck, L. B. Combinatorial receptor codes for odors. *Cell* **96**, 713–723 (1999).
15. Wilson, R. I. & Mainen, Z. F. Early events in olfactory processing. *Annu. Rev. Neurosci.* **29**, 163–201 (2006).
16. Luskin, M. B. & Price, J. L. The distribution of axon collaterals from the olfactory bulb and the nucleus of the horizontal limb of the diagonal band to the olfactory cortex, demonstrated by double retrograde labeling techniques. *J. Comp. Neurol.* **209**, 249–263 (1982).
17. Scott, J. W., McBride, R. L. & Schneider, S. P. The organization of projections from the olfactory bulb to the piriform cortex and olfactory tubercle in the rat. *J. Comp. Neurol.* **194**, 519–534 (1980).
18. Shipley, M. T. & Adamek, G. D. The connections of the mouse olfactory bulb: a study using orthograde and retrograde transport of wheat germ agglutinin conjugated to horseradish peroxidase. *Brain Res. Bull.* **12**, 669–688 (1984).
19. Nevelle, K. R. & Haberly, L. B. in *The Synaptic Organization of the Brain* (ed. Shepherd, G. M.) (Oxford Univ. Press, 2004).
20. Feinstein, P. & Mombaerts, P. A contextual model for axonal sorting into glomeruli in the mouse olfactory system. *Cell* **117**, 817–831 (2004).
21. Schaefer, M. L., Finger, T. E. & Restrepo, D. Variability of position of the P2 glomerulus within a map of the mouse olfactory bulb. *J. Comp. Neurol.* **436**, 351–362 (2001).
22. Brunjes, P. C., Illig, K. R. & Meyer, E. A. A field guide to the anterior olfactory nucleus (cortex). *Brain Res. Rev.* **50**, 305–335 (2005).
23. Yan, Z. *et al.* Precise circuitry links bilaterally symmetric olfactory maps. *Neuron* **58**, 613–624 (2008).
24. Schoenfeld, T. A. & Macrides, F. Topographic organization of connections between the main olfactory bulb and pars externa of the anterior olfactory nucleus in the hamster. *J. Comp. Neurol.* **227**, 121–135 (1984).
25. Scott, J. W., Ranier, E. C., Pemberton, J. L., Orona, E. & Mouradian, L. E. Pattern of rat olfactory bulb mitral and tufted cell connections to the anterior olfactory nucleus pars externa. *J. Comp. Neurol.* **242**, 415–424 (1985).
26. Miyamichi, K., Serizawa, S., Kimura, H. M. & Sakano, H. Continuous and overlapping expression domains of odorant receptor genes in the olfactory epithelium determine the dorsal/ventral positioning of glomeruli in the olfactory bulb. *J. Neurosci.* **25**, 3586–3592 (2005).
27. Yoshihara, Y. *et al.* OCAM: a new member of the neural cell adhesion molecule family related to zone-to-zone projection of olfactory and vomeronasal axons. *J. Neurosci.* **17**, 5830–5842 (1997).
28. Swanson, L. W. & Petrovich, G. D. What is the amygdala? *Trends Neurosci.* **21**, 323–331 (1998).
29. Scalia, F. & Winans, S. S. The differential projections of the olfactory bulb and accessory olfactory bulb in mammals. *J. Comp. Neurol.* **161**, 31–55 (1975).
30. Poo, C. & Isaacson, J. S. Odor representations in olfactory cortex: "sparse" coding, global inhibition, and oscillations. *Neuron* **62**, 850–861 (2009).
31. Stettler, D. D. & Axel, R. Representations of odor in the piriform cortex. *Neuron* **63**, 854–864 (2009).
32. Lei, H., Mooney, R. & Katz, L. C. Synaptic integration of olfactory information in mouse anterior olfactory nucleus. *J. Neurosci.* **26**, 12023–12032 (2006).
33. Franks, K. M. & Isaacson, J. S. Strong single-fiber sensory inputs to olfactory cortex: implications for olfactory coding. *Neuron* **49**, 357–363 (2006).
34. Stokes, C. C. & Isaacson, J. S. From dendrite to soma: dynamic routing of inhibition by complementary interneuron microcircuits in olfactory cortex. *Neuron* **67**, 452–465 (2010).
35. Apicella, A., Yuan, Q., Scanziani, M. & Isaacson, J. S. Pyramidal cells in piriform cortex receive convergent input from distinct olfactory bulb glomeruli. *J. Neurosci.* **30**, 14255–14260 (2010).
36. Suzuki, N. & Bekkers, J. M. Inhibitory neurons in the anterior piriform cortex of the mouse: classification using molecular markers. *J. Comp. Neurol.* **518**, 1670–1687 (2010).
37. Christie, J. M. *et al.* Connexin36 mediates spike synchrony in olfactory bulb glomeruli. *Neuron* **46**, 761–772 (2005).
38. Haberly, L. B. Parallel-distributed processing in olfactory cortex: new insights from morphological and physiological analysis of neuronal circuitry. *Chem. Senses* **26**, 551–576 (2001).
39. Kobayakawa, K. *et al.* Innate versus learned odour processing in the mouse olfactory bulb. *Nature* **450**, 503–508 (2007).
40. Marin, E. C., Jefferis, G. S., Komiyama, T., Zhu, H. & Luo, L. Representation of the glomerular olfactory map in the *Drosophila* brain. *Cell* **109**, 243–255 (2002).
41. Wong, A. M., Wang, J. W. & Axel, R. Spatial representation of the glomerular map in the *Drosophila* protocerebrum. *Cell* **109**, 229–241 (2002).
42. Jefferis, G. S. *et al.* Comprehensive maps of *Drosophila* higher olfactory centers: spatially segregated fruit and pheromone representation. *Cell* **128**, 1187–1203 (2007).
43. Davis, R. L. Olfactory memory formation in *Drosophila*: from molecular to systems neuroscience. *Annu. Rev. Neurosci.* **28**, 275–302 (2005).
44. Murthy, M., Fiete, I. & Laurent, G. Testing odor response stereotypy in the *Drosophila* mushroom body. *Neuron* **59**, 1009–1023 (2008).
45. Marshel, J. H., Mori, T., Nielsen, K. J. & Callaway, E. M. Targeting single neuronal networks for gene expression and cell labeling *in vivo*. *Neuron* **67**, 562–574 (2010).
46. Stepien, A. E., Tripodi, M. & Arber, S. Monosynaptic rabies virus reveals premotor network organization and synaptic specificity of cholinergic partition cells. *Neuron* **68**, 456–472 (2010).
47. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neurosci.* **13**, 133–140 (2010).
48. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).

METHODS

Generation of *CAG-stop-tTA2* Mice. The *tTA2* transactivator gene[49] was placed after the *CAG* promoter of plasmid *pCA-HZ2* (ref. 50) using a polymerase chain reaction (PCR)-based cloning method. A neomycin resistance (*neo^r*) gene and a transcriptional stop signal[51] were flanked by *loxP* sites to create a *loxP-neo^r-stop-loxP* cassette. This cassette was then introduced between the *CAG* promoter and *tTA2* using PCR-based cloning. An EcoRI fragment obtained from *ETLpA-/LTNL*[13], which contains the *IRES-tau-lacZ* cassette, was introduced after the *tTA2* coding sequence. The resulting cassette (*CAG-stop-tTA2-IRES-tau-lacZ*) was cloned into *pBT264* to flank the cassette with two copies of a ~250-bp β-globin *HS4* insulator sequence[52] on each side. *pBT264* (*pii-TRE-tdTomato-3Myc-ii*) was generated by inserting PCR-amplified copies of ~250-bp-long core insulator fragments (*i*) from the chicken β-globin HS4 insulator on each side of *TRE-tdTomato-3Myc* in *pBT239*. The insulator fragments were amplified from *pJC13-1* (ref. 53). The final construct, *pKM1* (*pii-CAG-stop-tTA2-IRES-tau-lacZii*), was tested by transient co-transfection with *pBT264* into cultured HEK293 cells. When a Cre-encoding plasmid *pBT140* (cytomegalovirus (CMV) promoter driving nuclear localization signal-*Cre*) was further introduced into the same cell, strong tdTomato fluorescence was detected 72 h after transfection. *pKM2* was digested with restriction enzymes SwaI and AscI, the insert was gel-purified using Qiagen gel extraction kit and eluted into 10 mM Tris-HCl, pH 7.4, 0.1 mM EDTA. The purified and linearized DNA devoid of plasmid backbone was used for mouse transgenesis via standard pronuclear injection procedure. Founders were screened by PCR primers to detect the *neo^r* gene. Four independent transgenic lines were established. They were crossed with mice containing *β-actin-CreER*[7] and *TRE-Bi-SG-T* reporter[53] to screen for functional *CAG-stop-tTA2* transgenes. Mice containing all three transgenes were injected with 1 mg of tamoxifen in corn oil at postnatal day (PD)10, and brains were collected at PD21 for the analysis. Two lines showed broad tdTomato fluorescence throughout the brain. One line (containing 2–3 copies of the transgene based on Southern blotting) was used exclusively in this study.

Virus preparations. All viral procedures followed the Biosafety Guidelines approved by the Stanford University Administrative Panel on Laboratory Animal Care (A-PLAC) and Administrative Panel of Biosafety (APB). To make the AAV containing the *TRE-HTG* cassette, which encodes histone–GFP, TVA and G linked by the 2A 'self-cleaving' peptides, the *HTG* cassette obtained from *pBOB-synP-HTB* (I.W. and E.M.C., unpublished plasmid) was placed after the *TRE-Tight* promoter in *pTRE-Tight* (Clontech), and then the entire construct was subcloned into the pAAV vector (Stratagene). Recombinant AAV serotype 2 was produced using the pAAV helper free kit (Stratagene) according to the manufacturer's instructions. AAV was also produced commercially by the Gene Therapy Center of the University of North Carolina. The AAV titre was estimated to be ~4×10^{12} viral particles ml^{-1} based on serial dilution and blot hybridization analysis. Pseudotyped *ΔG* rabies virus was prepared as previously described[3,54]. The pseudotyped rabies virus titre was estimated to be ~5×10^8 infectious particles per ml based on the infections of cell line 293-TVA800 by serially diluted virus stocks.

Animal surgery. All animal procedures followed animal care guidelines approved by A-PLAC. To activate Cre in animals carrying a *CreER* transgene, we injected intraperitoneally 0.1–1 mg of tamoxifen (Sigma) dissolved in corn oil into mice around PD10. For trans-synaptic labelling, 0.1–0.3 μl of AAV-*TRE-HTG* was injected into brain at PD21 by using a stereotactic apparatus (KOPF). During surgery, animals were anaesthetized with 65 mg kg^{-1} ketamine and 13 mg kg^{-1} xylazine (Ben Venue Laboratories). For motor cortex injections, the needle was placed 1.5 mm anterior and 1.5 mm lateral from the Bregma, and 0.4 mm from the brain surface. For olfactory cortex injections, see Supplementary Fig. 1b for the stereotactic parameters. After recovery, animals were housed in regular 12 h dark/light cycles with food and water *ad libitum*. Two weeks later, 0.3 μl of pseudotyped rabies virus (*ΔG-mCherry*+EnvA) was injected into the same brain location under anaesthesia. After recovery, animals were housed in a biosafety room for 7 days to allow rabies virus to infect, trans-synaptically spread and express sufficient amount of mCherry to label presynaptic cells. All animals were healthy and their brain structures were normal 7 days after rabies virus infection, confirming non-pathogenicity of *ΔG* mutant rabies virus.

Tissue processing. Brain tissue was processed according to previously described procedures[55]. To set the common coronal plane among different animals, the cerebellum was cut off and the brain was embedded in the Optimum Cutting Temperature (OCT) compound (Tissue-Tek) with the cut surface facing the bottom of the mould. The brain was adjusted to ensure that the left–right axis was parallel to the section plane. Neither mCherry nor histone–GFP required immunostaining for visualization. In some cases, brain sections were immunostained for better signal preservation according to previously published methods[56] using the following antibodies: chicken anti-GFP (1:500; Aves Labs), rabbit anti-DsRed (1:1,000; Clontech), donkey anti-chicken fluorescein isothiocyanate

(FITC) and donkey anti-rabbit Cy3 (1:200; Jackson ImmunoResearch). In most trans-synaptic labelling experiments starting from the olfactory cortex, every one of four sections of the olfactory bulb was immunostained by the free-floating method with goat anti-OCAM (1:100; R&D Systems) and donkey anti-goat Alexa488 (Invitrogen) to label OCAM$^+$ olfactory receptor neuron axons. For immunostaining against GABA, 60-μm free-floating coronal sections were treated with rabbit anti-GABA (1:2,000 in PBS with 0.3% Triton-X100; Sigma) for 48 h. GABA$^+$ cells were visualized with donkey anti-rabbit Cy3 (1:200; Jackson ImmunoResearch). Sections were imaged with a Nikon CCD camera by using a 10× objective or by 1-μm optical sectioning using confocal microscopy (Zeiss 510).

3D reconstruction. To compare distribution of labelled glomeruli (olfactory bulb) and starter cells (AON) across different samples, we needed to map them in a common 3D reference frame. To do this, we first saved manual annotations carried out in Adobe Illustrator in a scalable vector graphics (SVG) format. The SVG file saved all the annotations as an extensible markup language (XML) file describing the ellipses and contours (defined later), making it feasible to accurately parse the information by MATLAB scripts. In the olfactory bulb, we represented all glomeruli as ellipses. We used the centre of mass for each ellipse to define a single point, and calculated the centre of mass of all the points to define the centre of each slice. For the AON, we defined the contour as the boundary between layer I and layer II, which can be clearly distinguished by differences in the density of 4′,6-diamidino-2-phenylindole (DAPI) staining. To define the centre of mass for each contour, we replaced it with a dense series of points and used these points to calculate the centre of mass. Now, each slice is represented by a series of points and the centre of mass contained within an SVG file. To assemble the slices represented by SVG files into a 3D shape, we first aligned the centre of mass for each slice to that of the previous slice to form the cylindrical (z-)axis. Then, we refined the alignment by sequentially applying the iterative closest points (ICP) algorithm[56], which can identify the local rotation and translation parameters for each slice to maximize the overlap with the previous slice. Once we had aligned all the slices in a sample to generate a 3D shape, we needed to identify an orientation for the polar axis that could be most reliably identified in different 3D reconstructions. As the olfactory bulb is ellipsoidal, the principle component analysis (PCA) can reliably find a plane that contains the z-axis and intersects the 3D shape to maximize the surface of the intersection (plane *m*). We then defined the polar axis to be contained within the plane *m*, perpendicular to the z-axis, and pointing in the dorsal direction. For the AON, we approximated the contours of the most posterior slide of the AON as a triangle and calculated the rotation around the z-axis that minimizes the distance of the three vertices to those of a standard AON sample. We applied the same rotation to the whole 3D shape. To define the orientation of the polar axis, we used the side of the triangle that connects two of its medial vertices and points in the dorsal direction. Then we defined the polar axis as the line that is parallel to it and that intersects the z-axis. Finally, we calculated the volume occupied by each shape and applied a uniform scaling factor to account for different sizes of the anatomical structures in different animals.

All the steps explained earlier were implemented in MATLAB, which ran automatically without human intervention to avoid biasing the registration results. Once we had registered each shape, we used a standard algorithm to extract surfaces from two-dimensional (2D) contours[57] to transform the point cloud into a triangulated mesh that could be saved in the visualization toolkit (VTK)[58] format for visualization and analysis purposes.

We used the following landmarks to map starter cells in the amygdala and piriform cortex (Fig. 3): appearance of the olfactory tubercle (Fig. 3, I); end of the olfactory tubercle (Fig. 3, II); appearance of the hippocampus (Fig. 3, III); and appearance of the dentate gyrus of the hippocampus, on the ventral edge of the cortex (Fig. 3, IV).

Quantitative analyses. For each tracing experiment where we analysed the distribution of labelled glomeruli along the dorsal–ventral axis using mean θ'_{OB} (Fig. 3e), we generated a corresponding random distribution of simulated mean θ'_{OB} (mean $^{sim}\theta'_{OB}$) from *M* glomeruli, where *M* is the number of labelled glomeruli in the injection. To generate this random distribution for each experiment, we randomly selected *M* glomeruli from a given 3D reconstruction model (generated from that injection) and calculated the mean $^{sim}\theta'_{OB}$ value for those randomly selected *M* glomeruli to get the mean $^{sim}\theta'_1$. We then repeated the same simulation 50,000 times to obtain mean $^{sim}\theta'_2$,..., mean $^{sim}\theta'_{50,000}$, and therefore to obtain the range of mean $^{sim}\theta'$ for *M* glomeruli that are randomly distributed throughout the olfactory bulb. Once we obtained distributions for mean $^{sim}\theta'_{OB}$ that corresponded to each injection, we compared the mean $^{sim}\theta'_{OB}$ distribution with the experimental mean θ'_{OB}. If the value for the experimental mean θ'_{OB} was outside of the 95% of the mean $^{sim}\theta'_{OB}$ distribution, we considered the glomerular distribution to be non-random for that sample.

Multiple regression analysis (Fig. 4b) was conducted by using Excel (Microsoft). Data from every experiment in Supplementary Table 1 ($n = 8$ for the AON, $n = 10$ for the piriform cortex using *actin-CreER*) was used for the left part of the Fig. 4b. Data from seven experiments obtained from *GAD2-CreER* in the anterior piriform cortex were used in Fig. 4b, right. The number of labelled mitral cells in the olfactory bulb was set as a dependent variable, Y, and the number of starter cells in layer k ($k = $ I, II, III) was set as an independent variable, X_k. The constant was set to zero. Excel then calculated the values of coefficients A_k (shown by red crosses in Fig. 4b) and 95% confidence intervals of A_k based on the student's t-test (shown by grey bars in Fig. 4b). R^2 values for these multiple regression assays were: 0.98 for the AON; 0.96 for the piriform cortex (*actin-CreER*); and 0.97 for the piriform cortex (*GAD2-CreER*) data sets.

To estimate the number of dually labelled glomeruli originating from single starter cells (data Ds in Fig. 4c) in our experimental data, we first simulated a hypothetical number of dually labelled glomeruli originating from single starter cells (Ds) and two independent starter cells (Dt) according to the null hypothesis that mitral cells connect randomly with postsynaptic targets. This situation can be modelled by 'balls and bins': there are 2,000 bins (a bin represents a single glomerulus) and N balls (a ball represents a single trans-synaptic labelling event). N balls were randomly thrown into 2,000 bins, and the number of bins that received more than one ball (that is, glomeruli labelled more than once) was counted. To distinguish Ds from Dt, we further introduced n different colours to the balls, where each colour represented an individual starter cell in the cortex. We assumed that an equal number of balls (N/n) were labelled with n different colours. Each ball was randomly thrown into one of 2,000 bins, and the number of bins containing more than one ball was counted. We separately counted the bins with more than one ball of an identical colour (representing Ds) and the bins with more than one ball of different colours (representing Dt). We fixed the number of bins (glomeruli) to be 2,000, while N and n corresponded to the number of labelled mitral cells and the number of starter cells, respectively, in each experiment. We repeated this simulation 100,000 times for each set of N and n to obtain the simulated distribution of Ds and Dt (we call these 'random Ds' and

'random Dt'). To estimate the Ds components in experimental data (data Ds), we assumed that individual starter cells contributed independently to the labelling (random Dt = data Dt). On the basis of the equation: D (number of observed dually labelled glomeruli) = Ds + Dt, we estimated the data Ds distribution by subtracting the random Dt from observed D (Fig. 4c). Then we determined if there was a significant difference in the distribution of data Ds and random Ds. We considered two distributions to be significantly different if the probability of data Ds > random Ds or data Ds < random Ds exceeded 0.95 (shown by asterisks in Fig. 4c). To accurately count dually labelled glomeruli, samples with more than 200 labelled mitral cells were excluded from this analysis.

49. Urlinger, S. *et al.* Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity. *Proc. Natl Acad. Sci. USA* **97,** 7963–7968 (2000).
50. Zong, H., Espinosa, J. S., Su, H. H., Muzumdar, M. D. & Luo, L. Mosaic analysis with double markers in mice. *Cell* **121,** 479–492 (2005).
51. Muzumdar, M. D., Luo, L. & Zong, H. Modeling sporadic loss of heterozygosity in mice by using mosaic analysis with double markers (MADM). *Proc. Natl Acad. Sci. USA* **104,** 4495–4500 (2007).
52. Chung, J. H., Whiteley, M. & Felsenfeld, G. A 5′ element of the chicken β-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila. Cell* **74,** 505–514 (1993).
53. Li, L. *et al.* Visualizing the distribution of synapses from individual neurons in the mouse brain. *PLoS ONE* **5,** e11503 (2010).
54. Wickersham, I. R., Sullivan, H. A. & Seung, H. S. Production of glycoprotein-deleted rabies viruses for monosynaptic tracing and high-level gene expression in neurons. *Nature Protocols* **5,** 595–606 (2010).
55. Espinosa, J. S., Wheeler, D. G., Tsien, R. W. & Luo, L. Uncoupling dendrite growth and patterning: single-cell knockout analysis of NMDA receptor 2B. *Neuron* **62,** 205–217 (2009).
56. Zhang, Z. Iterative point matching for registration of free-form curves and surfaces. *Int. J. Comput. Vis.* **13,** 119–152 (1994).
57. Cong, G. & Parvin, B. Surface reconstruction from sparse fringe contours. *Proc. 4th IEEE Workshop Applicat. Comput. Vis.* **140** (1998).
58. Schroeder, W., Martin, K. & Lorensen, B. *The Visualization Toolkit: An Object-Oriented Approach to 3-D Graphics* (Prentice Hall, 1997).

LETTER

Magneto–thermal convection in solar prominences

Thomas Berger[1], Paola Testa[2], Andrew Hillier[3], Paul Boerner[1], Boon Chye Low[4], Kazunari Shibata[3], Carolus Schrijver[1], Ted Tarbell[1] & Alan Title[1]

Coronal cavities are large low-density regions formed by hemispheric-scale magnetic flux ropes suspended in the Sun's outer atmosphere[1]. They evolve over time, eventually erupting as the dark cores of coronal mass ejections[2,3]. Although coronal mass ejections are common and can significantly affect planetary magnetospheres, the mechanisms by which cavities evolve to an eruptive state remain poorly understood. Recent optical observations[4] of high-latitude 'polar crown' prominences within coronal cavities reveal dark, low-density[5] 'bubbles' that undergo Rayleigh–Taylor instabilities[6,7] to form dark plumes rising into overlying coronal cavities. These observations offered a possible mechanism for coronal cavity evolution, although the nature of the bubbles, particularly their buoyancy, was hitherto unclear. Here we report simultaneous optical and extreme-ultraviolet observations of polar crown prominences that show that these bubbles contain plasma at temperatures in the range $(2.5–12) \times 10^5$ kelvin, which is 25–120 times hotter than the overlying prominence. This identifies a source of

the buoyancy, and suggests that the coronal cavity–prominence system supports a novel form of magneto-thermal convection in the solar atmosphere, challenging current hydromagnetic concepts of prominences and their relation to coronal cavities.

On 22 June 2010, the Hinode[8] Solar Optical Telescope[9] (SOT) observed a quiescent prominence at the northwest solar limb (56° N, 90° W). SOT obtained optical images in the Ca II H-line at 396.8 nm, and images and line-of-sight velocity 'dopplergrams' in the Hα 656.3 nm line, both with ~150 km spatial resolution. These images reveal the fine-scale structure and dynamics of the low-temperature ($\sim 7–10 \times 10^3$ K) plasma in the prominence. The Hinode X-ray Telescope[10] obtained co-temporal images in X-rays that showed the high-temperature ($10^6–10^7$ K) areas of the associated coronal cavity structure. Simultaneously, the Solar Dynamics Observatory Atmospheric Imaging Assembly[11] (AIA) instrument acquired full-Sun images in 10 visible, ultraviolet and extreme-ultraviolet wavelengths
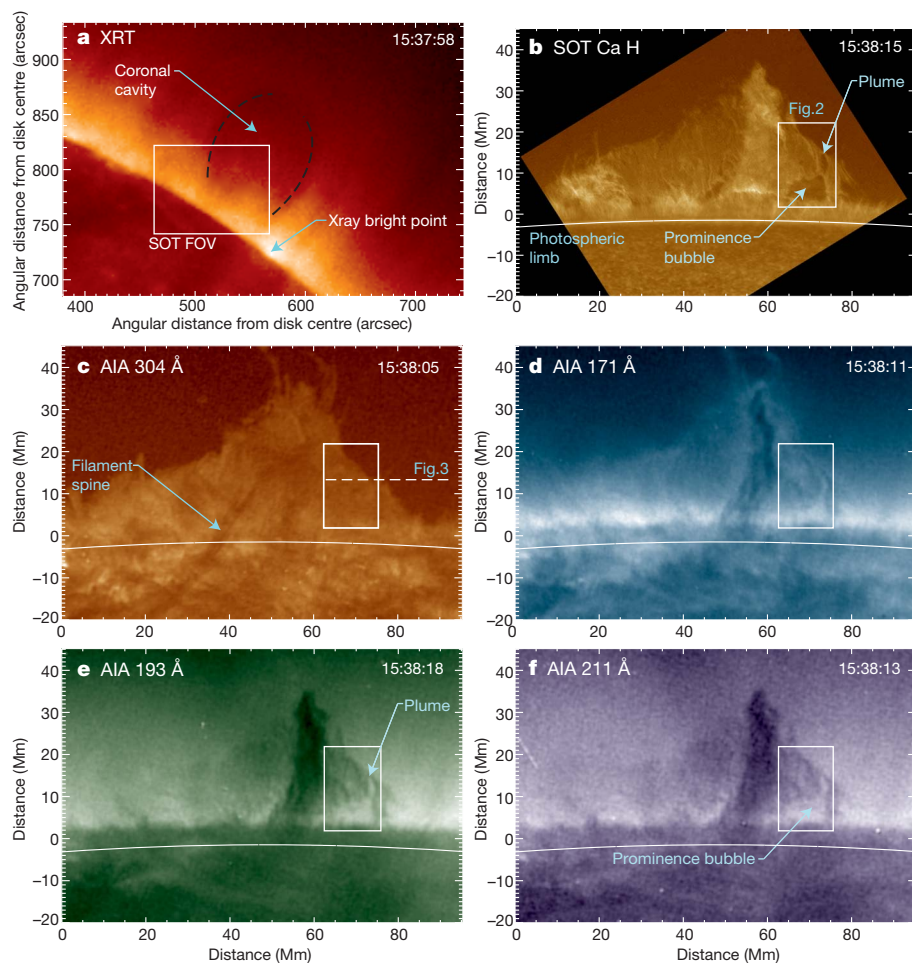
Figure 1 | Quiescent prominence observed on the northwest solar limb on 22 June 2010. a, Hinode X-Ray Telescope image of the region surrounding the prominence. The coronal cavity is demarcated by the black dashed curves. The prominence is transparent to X-rays and is not seen in this image. The SOT field of view (FOV) is indicated by the white box. **b**, SOT image in the Ca II H-line. The bubble on the right side of the prominence is noted, as is a large rising plume. The white box outlines the FOV shown in Fig. 2. The photospheric limb is rotated to be horizontal, and is demarcated by the white curve in this and all subsequent images. **c**, AIA 304-Å channel image. Note that the bubble is not visible in this passband, indicating that optically thick emission is mostly from the sheath-like prominence–corona transition region surrounding the structure seen in **b**. The plume is seen in the 304-Å movie as a rising bump along the right edge of the prominence. The dashed horizontal line indicates a line along which prominence, plume, and background coronal emission are sampled (Fig. 3). **d**, AIA 171-Å channel image. The prominence shows strong absorption along the spine region and weak emission from the surrounding regions. **e**, AIA 193-Å channel image. This and the subsequent 211-Å channel image show strong absorption from neutral H and neutral and ionized He in all regions of the prominence. **f**, AIA 211-Å channel image. Note that the bubble and plume show up clearly in emission relative to the dark prominence. The darkened region above 30 Mm indicates the lower regions of the associated coronal cavity. In **b–f**, y axes have been set to zero at the approximate location of the photospheric limb. See Supplementary Movies 1–8 for time series of all SOT and AIA images used in this work. Time (UT) is shown at top right in all panels.

[1]Lockheed Martin Advanced Technology Center, Solar and Astrophysics Laboratory, O/ADBS B/252, 3251 Hanover Street, Palo Alto, California 94304, USA. [2]Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA. [3]Kwasan and Hida Observatories, Graduate School of Science, Kyoto University, Kyoto 607-8471, Japan. [4]High Altitude Observatory, National Center for Atmospheric Research, PO Box 3000, Boulder, Colorado 80307, USA.

with a spatial resolution of approximately 870–1,230 km. We use the AIA spectral channels at (Å) 131, 171, 193, 211 and 304 to analyse coronal plasma emission in the prominence and its surroundings. The SOT Ca H-line image shows a central bright column about 37 Mm high in the plane of the sky with triangular regions on either side (Fig. 1). The width of the prominence, measured along the limb, is 77 Mm. The AIA 304-Å channel shows optically thick emission from plasma at 8×10^4 K in the prominence as well as absorption in the dark overlying 'spine' filaments. The other AIA images show optically thin emission, or its absorption by hydrogen and helium, from $\sim 10^6$ K plasma.

SOT movies reveal a dark semi-circular dynamic void below the right triangular region. The void is a prominence bubble approximately 9.5 ± 0.5 Mm wide, varying from 7.5 to 11.0 (± 0.5) Mm in height over 3 h. An elongation in the upper right spawns a plume 1.4 ± 0.5 Mm wide that follows a curved trajectory with an average speed (on the plane of the sky) of 22.8 ± 5 km s^{-1}. The curved trajectory implies that the plume follows magnetic field lines. The plume is identifiable in the images for approximately 530 s. The plume boundary is red-shifted by 4.1 ± 1 km s^{-1}, implying that the velocity is 23.2 ± 5 km s^{-1} at 10° to the plane of the sky. In the AIA EUV images, both the bubble and plume show up in persistent emission relative to the darker prominence material (Fig. 2).

The observation that the bubble and plume are bright in high-temperature EUV images indicates the possibility of significant heating low in the prominence. To analyse the bubble and plume EUV emission in more detail, we measure the contrast of the structures relative to 'background' coronal emission outside the prominence. Contrast is defined by $(\langle I \rangle / I_B) - 1$, with $\langle I \rangle$ the spatially averaged intensity in the bubble or plume, and I_B the average background level. Bubble contrast in the 304-Å channel is positive throughout the measurements but the hotter EUV channels exhibit negative contrast (Fig. 3). These results are not conclusive, however, as He II 304-Å emission is optically thick and I_B at the bubble height is contaminated by limb brightening and a nearby X-ray bright point. However the plume is located above the limb brightening and yields a more reliable contrast measurement. We measure positive plume contrast in both the 304-Å and 171-Å channels, the latter decreasing over time as the plume rises and cools. Positive contrast at 171 Å indicates that the plume contains plasma hotter than the ambient corona. As the plume originates from the bubble, the bubble must be at least as hot.

To estimate the temperature, we use the two-filter ratio technique with the known temperature response of the AIA spectral channels. Ratios of bubble and plume emission in the 131-, 171-, 193- and 211-Å channels are formed and plotted over the equivalent telescope response ratios. Owing to the wide and sometimes double-peaked AIA response curves, the measured ratios intersect the response curves at two solutions: $\log T \approx 5.4$ (2.5×10^5 K) and $\log T \approx 6.1$ (1.2×10^6 K). Although bifurcated, the solutions indicate that the minimum temperature in the structures is at least 2.5×10^5 K. Moreover, observed radiative cooling times for the plume, as well as estimates of heating rates for the bubble, suggest that the hotter solution is more likely. Such temperatures have been detected in the thin 'prominence–corona transition region'[12], but not in large-scale buoyant structures below prominences.

A larger bubble in a quiescent prominence was observed by AIA on 2 July 2010 at 46° S, 90° E. The 304-Å channel shows a 22-Mm-wide bubble developing over 2.5 h (Fig. 4). Large bubbles have been seen in SOT and rare ground-based[13,14] observations, but no explanation for their origin has been given. Unlike the bubble observed on 22 June 2010 that rises ~ 10 Mm and then stagnates, this larger bubble continually grows until it reaches a peak height 43.5 ± 0.5 Mm. The top boundary then undergoes Rayleigh–Taylor instabilities and the bubble collapses. The 171-Å channel shows a hot compact region of plasma emerging from the chromosphere and propagating upwards to the right of the bubble centre. The region rises with a speed of 15.3 ± 1 km s^{-1}, pushing the cooler plasma upwards to form the
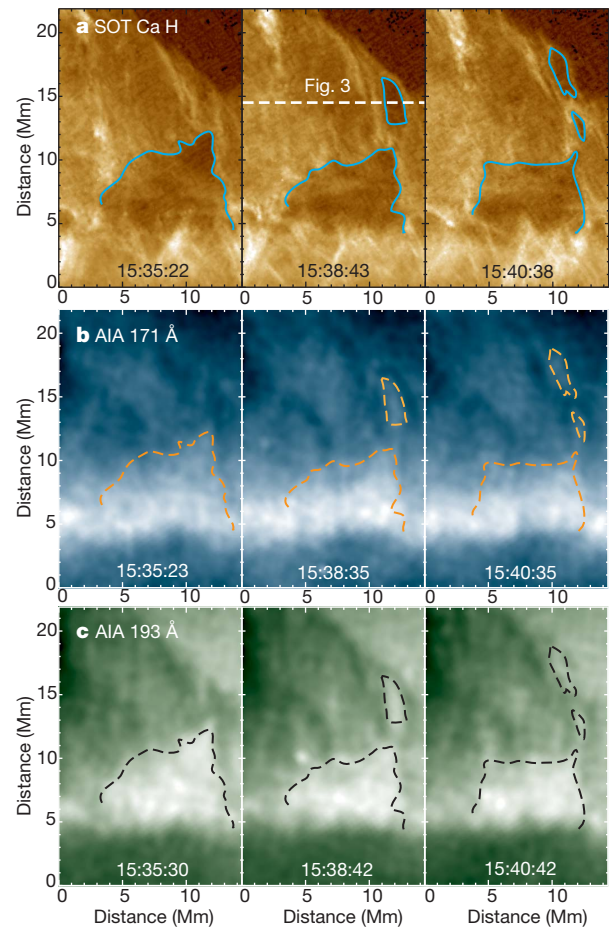
Figure 2 | Time series of images taken during the large plume generation event in the prominence observed on 22 June 2010. All images have the solar limb rotated to horizontal. Time (in UT) is shown at bottom in all panels. **a**, SOT Ca II H-line images. The bubble and plume are dark in this spectral channel and are easily manually outlined to define the structures (solid blue lines). The horizontal dashed line in the 15:38:43 UT image denotes the spatial sample line along which plume emission is measured relative to the prominence and coronal background. **b**, Co-temporal AIA 171-Å channel series with the bubble and plume definition contours from **a**. **c**, Co-temporal AIA 193 channel series with the same contours. In both EUV channels (**b** and **c**) the bubble and plume show up as bright structures relative to the surrounding prominence. The spatial correlation of the EUV brightening with the outlines of the bubble and plume in the SOT images is precise to <1 Mm. See Supplementary Movie 9 for the full time series of these data.

bubble. The average filter ratios (in Å) of 171/193, 171/211 and 193/211 in the compact region again define bifurcated temperature solutions of $\log T \approx 5.5$–5.7 and $\log T \approx 6.0$. However, the overlap of the ratios is significantly better in the $\log T \approx 6.0$ range and we conclude, as above, that the hotter solution is more likely.

The high temperature in prominence bubbles explains both their darkness in optical observations and their enhanced buoyancy: at $5.4 < \log T < 6.0$, hydrogen and calcium are ionized beyond the temperatures of Hα or Ca II H-line formation, and the thermal energy density in the bubble imparts significant buoyancy relative to an overlying prominence with a temperature of $\log T < 4$. The most likely heat sources are spontaneous current sheet formation[15] and Alfvèn wave dissipation[16] in the low-density bubble plasma. Using an established heat flux requirement[17], an electron number density of 10^{15} m^{-3}, and measured bubble sizes, the heating time is approximately 2×10^3 s, significantly less than the bubble lifetimes and compatible with observed growth rates.

The sharply defined boundary of the bubbles indicates a metastable state in which buoyancy is balanced by tension in the magnetic field
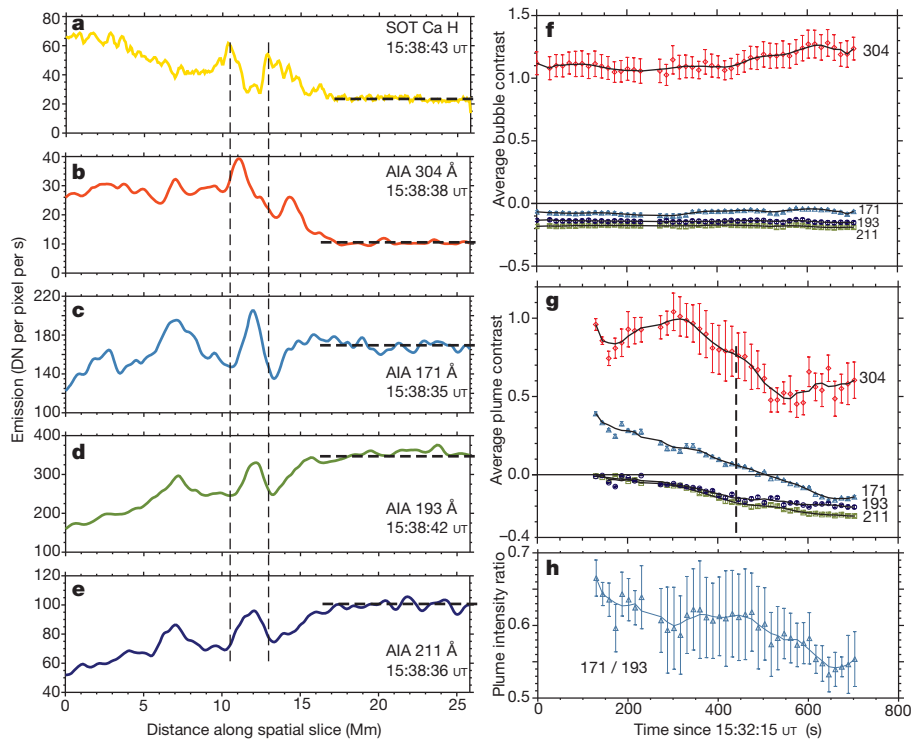
Figure 3 | Emission, contrast and AIA filter ratio analysis for the bubble and plume in the 22 June 2010 quiescent prominence. a–e, SOT Ca II H-line (**a**), AIA 304 Å (**b**), 171 Å (**c**), 193 Å (**d**) and 211 Å (**e**) channel emission traces taken along the dashed line shown in Fig. 1c (through the plume) at approximately 15:38:40 UT. This time corresponds to the middle column of images in Fig. 2. DN stands for Data Number, a generic count of digitized camera signal. The plume location is defined by the vertical dashed lines in **a**. The horizontal dashed lines indicate the background emission to the right of the prominence in all channels. Note that the AIA 304-Å and 171-Å emission levels are above background coronal levels at this time while the 193-Å and 211-Å emission levels are slightly below. The elevated background emission in the hotter EUV channels may be due to the nearby X-ray bright point visible in Fig. 1a. **f,** Bubble contrast as a function of time from 15:32:15 UT. Symbols denote times of individual images in each channel. The emission over the manually defined bubble region (see Fig. 2) is spatially averaged in each image

before comparison to background levels. The elevated EUV background levels noted above make the 171-Å, 193-Å and 211-Å contrast values negative for the entire time series. **g,** Plume contrast as a function of time since 15:32:15 UT. Symbols denote times of individual images in each channel. The emission is spatially averaged in each image before comparison to background. Here the 171-Å channel contrast is positive for the first ~500 s of the plume lifetime. The 193-Å and 211-Å contrast values are suppressed by the elevated background levels seen in **d** and **e**. Error bars in **f** and **g** are 1σ deviations on the mean contrast values; bars are smaller than the symbols in the 171-Å, 193-Å and 211-Å channels. **h,** AIA 171 Å/193 Å filter intensity ratio for the prominence plume as a function of time since 15:32:15 UT. Symbols denote times of 171-Å images; the 'nearest neighbour' 193-Å image is used in the ratio. Error bars denote uncertainty in the ratios based on the 1σ errors on the emission values in **f** and **g**. All other two-filter ratios used in the analysis show similar time profiles and error bar magnitudes.

of the overlying prominence, suppressing high-wavenumber perturbations and allowing the density inversion to remain coherent on multi-hour timescales. However, the bubbles eventually undergo Rayleigh–Taylor instability and either slowly deflate (as observed on 22 June) or rapidly collapse (as observed on 2 July). Numerical simulations of the lower solar atmosphere show that the Rayleigh–Taylor

instability is a natural consequence of density inversions in the gravitational field[18]. New prominence simulations motivated by this finding show that the instability initiates where the thermal perturbation is largest and/or the overlying field strength is weakest. Bubbles and plumes in prominences can thus be understood as magneto-thermal convective instabilities in the corona—distinct from supposed coronal
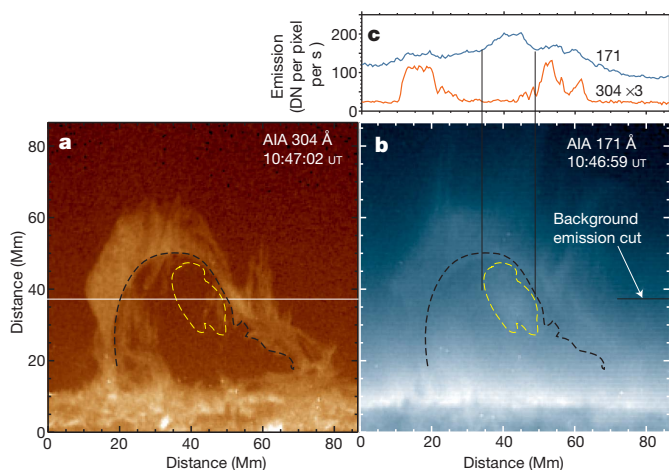
Figure 4 | Quiescent prominence observed on the southeast solar limb on 2 July 2010. a, AIA 304-Å channel image showing a large prominence bubble rising to a height of 43 Mm above the limb. The black dashed curve traces the edges of the bubble in the 304-Å channel; the yellow dashed curve traces the hot central region defined in the 171-Å channel image. The horizontal white line denotes the spatial sampling line used in **c**. **b,** AIA 171-Å channel image showing the bubble and hot rising structure in dashed contour lines. **c,** Spatial samples from the 304-Å channel (orange) and the 171-Å channel (blue), showing the emission in DN per pixel per second along the horizontal line in **a**. The black vertical lines show that the increased intensity in the 171-Å channel occurs at the location of the rising structure outlined by the dashed contour. The enhancement of the region over the background level (~100 DN) is approximately a factor of 2. The 304-Å channel plot has been multiplied by a factor of 3 for comparison purposes. See Supplementary Movie 10 for a time series of these data.

loop convection[19,20]—that transport hot plasma, and presumably magnetic flux, from chromospheric levels into the overlying coronal cavities.

We speculate that the origin of the bubbles is the emergence of twisted magnetic flux below prominences, as has been detected in SOT filament observations[21–23]. Models of flux emergence into magnetized atmospheres show that the maximum height reached is a sensitive function of the overlying magnetic field strength[24]. Using typical prominence characteristics in such models gives good agreement with the heights reached by the prominence bubbles shown here, but further measurements are required to confirm a flux emergence origin.

Doppler velocity measurements in coronal cavities[25] suggest that intermittent flows occur in these structures. Our finding shows that coronal-temperature plasma can be injected into coronal cavities from below, offering an explanation for the flows. Although the size of the bubbles is much less than that of the overlying coronal cavity, SOT observations show that bubbles can recur on timescales of 5–10 h in roughly the same location. This implies that quiescent prominence convection is a gradual but consistent process that could supply significant amounts of plasma, magnetic flux and helicity to the cavity over time. The finding supports theories of coronal mass ejections in which the gradual build-up of magnetic flux and helicity leads to the destabilization and eruption of coronal flux ropes[26,27].

1. Pneuman, G. W. Temperature-density structure in coronal helmets: the quiescent prominence and coronal cavity. *Astrophys. J.* **177,** 793–805 (1972).
2. Gosling, J. T., Hildner, E., MacQueen, R. M., Munro, R. H. & Poland, A. I. Mass ejections from the Sun: a view from Skylab. *J. Geophys. Res.* **79,** 4581–4587 (1974).
3. Zhang, M. & Low, B. C. The hydromagnetic nature of coronal mass ejections. *Annu. Rev. Astron. Astrophys.* **43,** 103–137 (2005).
4. Berger, T. E. *et al. Hinode* SOT observations of solar quiescent prominence dynamics. *Astrophys. J.* **676,** L89–L92 (2008).
5. Heinzel, P. *et al.* Hinode, TRACE, SOHO, and ground-based observations of a quiescent prominence. *Astrophys. J.* **686,** 1383–1396 (2008).
6. Berger, T. E. *et al.* Quiescent prominence dynamics observed with the Hinode Solar Optical Telescope. I. Turbulent upflow plumes. *Astrophys. J.* **716,** 1288–1307 (2010).
7. Ryutova, M., Berger, T., Frank, Z., Tarbell, T. & Title, A. Observation of plasma instabilities in quiescent prominences. *Sol. Phys.* **267,** 75–94 (2010).
8. Kosugi, T. *et al.* The *Hinode* (Solar-B) mission: an overview. *Sol. Phys.* **243,** 3–17 (2007).
9. Suematsu, Y. *et al.* The Solar Optical Telescope of Solar-B (*Hinode*): the optical telescope assembly. *Sol. Phys.* **249,** 197–220 (2008).
10. Golub, L. *et al.* The X-Ray Telescope (XRT) for the *Hinode* mission. *Sol. Phys.* **243,** 63–86 (2007).
11. Lemen, J. R. *et al.* The Atmospheric Imaging Assembly on the Solar Dynamics Observatory. *Sol. Phys.* (submitted) (2010).
12. Cirigliano, D., Vial, J.-C. & Rovira, M. Prominence corona transition region plasma diagnostics from SOHO observations. *Sol. Phys.* **223,** 95–118 (2004).
13. de Toma, G., Casini, R., Burkepile, J. T. & Low, B. C. Rise of a dark bubble through a quiescent prominence. *Astrophys. J.* **687,** L123–L126 (2008).
14. Stellmacher, G. & Wiehr, E. Observation of an instability in a ''quiescent'' prominence. *Astron. Astrophys.* **24,** 321–324 (1973).
15. Janse, M., Low, B. C. & Parker, E. N. Topological complexity and tangential discontinuities in magnetic fields. *Phys. Plasmas* **17,** 092901, doi:10.1063/1.3474943 (published online 8 September 2010).
16. Matsumoto, T. & Shibata, K. Nonlinear propagation of Alfvén waves driven by observed photospheric motions: application to the coronal heating and spicule formation. *Astrophys. J.* **710,** 1857–1867 (2010).
17. Withbroe, G. L. & Noyes, R. W. Mass and energy flow into the solar chromosphere and corona. *Annu. Rev. Astron. Astrophys.* **15,** 363–387 (1977).
18. Isobe, H., Miyagoshi, T., Shibata, K. & Yokoyama, T. Filamentary structure on the Sun from the magnetic Rayleigh-Taylor instability. *Nature* **434,** 478–481 (2005).
19. Foukal, P. Magnetic loops, downflows, and convection in the solar corona. *Astrophys. J.* **223,** 1046–1057 (1978).
20. Marsch, E., Tian, H., Sun, J., Curdt, W. & Wiegelmann, T. Plasma flows guided by strong magnetic fields in the solar corona. *Astrophys. J.* **685,** 1262–1269 (2008).
21. Okamoto, T. J. *et al.* Emergence of a helical flux rope under an active region prominence. *Astrophys. J.* **673,** L215–L218 (2008).
22. Lites, B. W. *et al.* Emergence of helical flux and the formation of an active region filament channel. *Astrophys. J.* **718,** 474–487 (2010).
23. Okamoto, T. J., Tsuneta, S. & Berger, T. E. A rising cool column as a signature of helical flux emergence and formation of prominence and coronal cavity. *Astrophys. J.* **719,** 583–590 (2010).
24. Shibata, K., Tajima, T., Steinolfson, R. S. & Matsumoto, R. Two-dimensional magnetohydrodynamic model of emerging magnetic flux in the solar atmosphere. *Astrophys. J.* **345,** 584–596 (1989).
25. Schmit, D. J. *et al.* Large-scale flows in prominence cavities. *Astrophys. J.* **700,** L96–L98 (2009).
26. Fan, Y. & Gibson, S. E. Onset of coronal mass ejections due to loss of confinement of coronal flux ropes. *Astrophys. J.* **668,** 1232–1245 (2007).
27. Zhang, M., Flyer, N. & Low, B. C. Magnetic field confinement in the corona: the role of magnetic helicity accumulation. *Astrophys. J.* **644,** 575–586 (2006).

LETTER

Universal spin transport in a strongly interacting Fermi gas

Ariel Sommer[1,2,3], Mark Ku[1,2,3], Giacomo Roati[4,5] & Martin W. Zwierlein[1,2,3]

Transport of fermions, particles with half-integer spin, is central to many fields of physics. Electron transport runs modern technology, defining states of matter such as superconductors and insulators, and electron spin is being explored as a new carrier of information[1]. Neutrino transport energizes supernova explosions following the collapse of a dying star[2], and hydrodynamic transport of the quark–gluon plasma governed the expansion of the early Universe[3]. However, our understanding of non-equilibrium dynamics in such strongly interacting fermionic matter is still limited. Ultracold gases of fermionic atoms realize a pristine model for such systems and can be studied in real time with the precision of atomic physics[4]. Even above the superfluid transition, such gases flow as an almost perfect fluid with very low viscosity when interactions are tuned to a scattering resonance[3,5–8]. In this hydrodynamic regime, collective density excitations are weakly damped[6,7]. Here we experimentally investigate spin excitations in a Fermi gas of ^{6}Li atoms, finding that, in contrast, they are maximally damped. A spin current is induced by spatially separating two spin components and observing their evolution in an external trapping potential. We demonstrate that interactions can be strong enough to reverse spin currents, with components of opposite spin reflecting off each other. Near equilibrium, we obtain the spin drag coefficient, the spin diffusivity and the spin susceptibility as a function of temperature on resonance and show that they obey universal laws at high temperatures. In the degenerate regime, the spin diffusivity approaches a value set by $\hbar/m$, the quantum limit of diffusion, where $\hbar$ is Planck's constant divided by 2π and m the atomic mass. For repulsive interactions, our measurements seem to exclude a metastable ferromagnetic state[9–11].

Understanding the transport of spin, as opposed to the transport of charge, is highly relevant to the novel field of spintronics[1]. Whereas charge currents are unaffected by electron–electron scattering owing to momentum conservation, spin currents are intrinsically damped owing to collisions between electrons of opposite spin, as their relative momentum is not conserved. This phenomenon is known as spin drag[12,13]. It is expected to contribute significantly to the damping of spin currents in doped semiconductors[14]. The random collision events also lead to spin diffusion—the tendency for spin currents to flow in such a way as to even out spatial gradients in the spin density—which has been studied in high-temperature superconductors[15] and in liquid ^{3}He–^{4}He mixtures[16,17].

Creating spin currents poses a major challenge in electronic systems, where mobile spins are scattered by their environment and by each other. However, in ultracold atomic gases, we have the freedom to first prepare an essentially non-interacting spin mixture, separate atoms spatially by using magnetic field gradients, and only then induce strong interactions. Past observations of spin currents in ultracold Fermi gases[18,19] were made in the weakly interacting regime. Here we access the regime near a Feshbach resonance[4], where interactions are as strong as allowed by quantum mechanics (the unitarity limit).

We measure spin transport properties, namely the spin drag coefficient Γ_{sd} and the spin diffusivity D_s, of a strongly interacting Fermi gas composed of an equal number of atoms in two different spin states. In the strongly interacting regime, spin drag is expected to reach a universal maximum value, and spin diffusion is expected to reach a universal minimum.

The universal behaviour of the spin transport coefficients of a Fermi gas can be estimated on general grounds. At the Feshbach resonance, the scattering cross-section σ between atoms of opposite spin is given by the square of the de Broglie wavelength. In the degenerate regime (that is, below the Fermi temperature T_F) $\sigma \approx 1/k_F^2$, where $k_F = (6\pi^2 n)^{1/3}$ is the Fermi wavevector and n is the density of atoms in each spin state. The mean free path between collisions is thus $l = 1/(n\sigma) \approx 1/k_F$, or about one interparticle spacing, which is the smallest possible mean free path in a gas. The average speed v of atoms is of the order of the Fermi velocity, $\hbar k_F/m$. In estimating the spin diffusivity $D_s \approx vl$, the density-dependent factors cancel, giving $D_s \approx \hbar/m$. This value for D_s represents a universal quantum limit to spin diffusivity in Fermi gases. Away from resonance, the scattering cross-section decreases, increasing D_s. For temperatures T much greater than T_F, the scattering cross-section will be given by the square of the thermal de Broglie wavelength, and thus decreases as $\sigma \propto 1/T$. The velocity, in turn, will increase as $v \propto \sqrt{T}$, causing D_s to increase as $D_s \propto T^{3/2}$. An analogous scaling argument applies to the viscosity[7,8]. Finally, in a degenerate Fermi gas, the average velocity will remain of the order of the Fermi velocity, but the effective scattering cross-section will scale as $\sigma \propto T^2$ owing to Pauli blocking, causing D_s to increase as T^{-2} as the temperature is lowered. For a Fermi gas, we thus expect the minimum D_s to occur near T_F, before Pauli blocking becomes effective. Correspondingly, the coefficient Γ_{sd} characterizing spin drag is expected to reach a universal maximum value, given by the Fermi rate $E_F/\hbar$, where $E_F = \hbar^2 k_F^2/2m$ is the Fermi energy.

In our experiment, we prepare an equal mixture of the two lowest hyperfine states ('spin up' and 'spin down') of fermionic ^{6}Li in a cylindrically symmetric atom trap[4,20,21]. The confinement along the axis of symmetry is harmonic, with frequency ω_z. We separate the two spin components along the axis of symmetry of the trap (see Methods Summary), and turn on strong interactions between unequal spins by quickly increasing the magnetic field to a Feshbach resonance located at 834 G. The confining potential of the trap forces the two clouds of opposite-spin atoms to propagate towards each other, establishing a spin current. Measurements are made by selectively imaging the two spin components.

Figure 1 shows the collision between the two spin domains on resonance. The clouds bounce off each other and essentially completely repel each other. Owing to the axial trapping potential, the clouds return after the collision, and we observe several oscillations in the displacement $d = \langle z_\uparrow \rangle - \langle z_\downarrow \rangle$, where $\langle z_{\uparrow(\downarrow)} \rangle$ is the centre of mass of the spin-up (spin-down) cloud. After the oscillations have decayed, the displacement decreases to zero monotonically, on a timescale of the

[1]Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [2]MIT-Harvard Center for Ultracold Atoms, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [4]INO-CNR, University of Florence, 50019 Sesto Fiorentino, Italy. [5]LENS, University of Florence, 50019 Sesto Fiorentino, Italy.

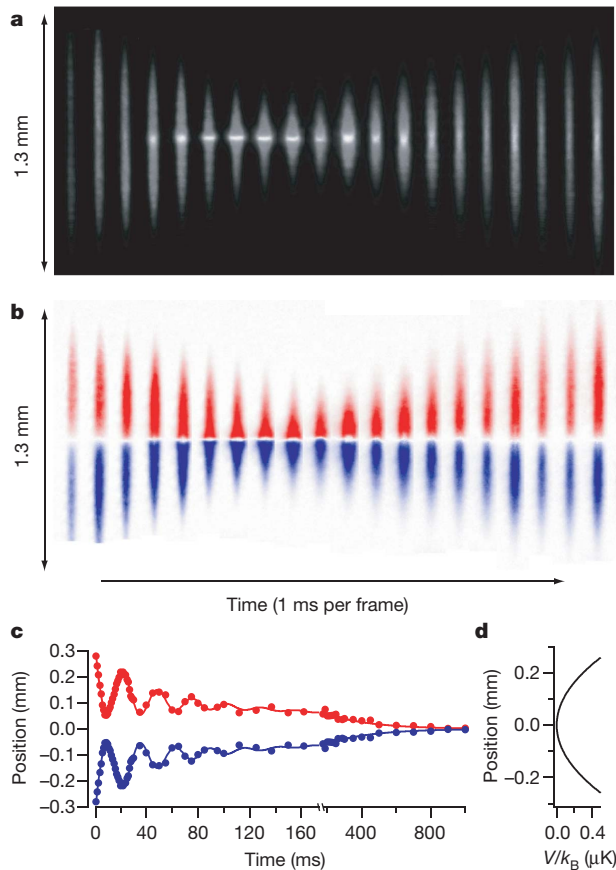**Figure 1 | Observation of spin current reversal in a resonant collision between two oppositely spin-polarized clouds of fermions. a, b,** Total column density (**a**) and the difference in column densities (**b**: red, spin up; blue, spin down) during the first 20 ms after the collision. The central column densities here are typically $7 \times 10^9 \, \text{cm}^{-2}$. Strong repulsion is observed that leads to a high-density interface. **c,** The centre of mass separation initially oscillates at 1.63(2) times the axial trap frequency of 22.8 Hz (see Supplementary Information) before decaying exponentially at later times. The initial atom number per spin state is 1.2×10^6, and the temperature 200 ms after the collision and later is $0.5T_F$, with T_F the Fermi temperature at the centre of each cloud. **d,** The trapping potential V is harmonic along the symmetry axis.

order of one second, which is an extremely long time compared to the trapping period (44 ms). The underlying explanation for spin current reversal and the slow relaxation can be found in the extremely short mean free path and the high collision rate between opposite-spin atoms at unitarity. According to the above estimate, the spin diffusivity is approximately $\hbar/m$, which for ^{6}Li is $(100 \, \mu\text{m})^2 \, \text{s}^{-1}$. The atom clouds in the experiment have a length of the order of $100 \, \mu\text{m}$, and it takes them of the order of a second to diffuse through each other. So we are indeed observing quantum-limited spin diffusion. The initial bounces will occur when the mean free path of a spin-up atom in the spin-down cloud is smaller than the spin-down cloud size, that is, when the mixture is hydrodynamic. Instead of quickly diffusing into the spin-down region, it is then more likely that the spin-up atom is scattered back into the spin-up region, where it can propagate ballistically.

After long evolution times, the oscillations shown in Fig. 1 have been damped out, and the displacement between the centres of mass is much smaller than the widths of the clouds. The relaxation dynamics can then be described by linear response theory, giving access to the spin transport coefficients. The spin drag coefficient Γ_{sd} is defined as the rate of momentum transfer between opposite-spin atoms[12,14], and is therefore related to the collision rate. From the Boltzmann transport equation, the relaxation of the displacement d near equilibrium follows the differential equation[22]

$$\Gamma_{sd}\dot{d} + \omega_z^2 d = 0$$

in the case of strongly overdamped motion realized here. Fitting an exponential with decay time τ to the displacement gives the spin drag coefficient of the trapped system as $\Gamma_{sd} = \omega_z^2 \tau$. In the deeply degenerate regime, the relationship between the measured and the microscopic spin drag coefficient might be affected by a weak enhancement of the effective mass[23] and the attractive interaction energy between the clouds[10,22,24].

The spin drag coefficient is found to be greatest on resonance, and thus spin conduction is slowest on resonance (see Supplementary Information). On resonance, Γ_{sd} in a homogeneous system must be given by a function of the reduced temperature T/T_F times the Fermi rate $E_F/\hbar$. At high temperatures, we expect the spin drag coefficient to obey a universal scaling $\Gamma_{sd} \propto n\sigma v \propto \frac{E_F}{\hbar}(T/T_F)^{-1/2}$. In Fig. 2 we show the spin drag coefficient as a function of T/T_F; Γ_{sd} is normalized by $E_F/\hbar$, where E_F and T_F are the local values at the centre of total mass. We observe $T^{-1/2}$ scaling for $T/T_F > 2$, finding $\Gamma_{sd} = 0.16(1)\frac{E_F}{\hbar}(T/T_F)^{-1/2}$. At lower temperatures, we observe a crossover from classical to non-classical behaviour as the spin drag coefficient reaches a maximum of approximately $0.1E_F/\hbar$ near the Fermi temperature. We interpret this saturation of the spin drag coefficient as a consequence of Fermi statistics and unitarity[4,5], as σ and v approach values determined by the Fermi wave-vector k_F. The spin drag coefficient is inversely proportional to the spin conductivity, which describes the spin current response to an external spin-dependent force. Near the Fermi temperature, the maximum spin drag coefficient corresponds to a minimum spin conductivity of the order of $k_F/\hbar$. This is the slowest spin conduction possible in three dimensions in the absence of localization.

At low temperatures, the spin drag coefficient decreases with decreasing temperature. Reduced spin drag at low temperatures is expected in Fermi liquids owing to Pauli blocking[11,18,22,24,25], and is also expected in one-dimensional Fermi gases[26]. In the case of collective density (rather than spin) excitations, it was shown that pairing correlations enhance the effective collision rate dramatically as the temperature is lowered[6]. The effect of pairing on the spin drag coefficient may be qualitatively different. In a simple picture, spin currents require the flow of unpaired atoms, whereas collective density excitations affect paired and unpaired atoms alike.

Comparing the relaxation rate to the gradient in spin density allows us to also measure the spin diffusivity D_s. At the centre of the trap, the spin current density J_s is given by the spin diffusion equation[27]
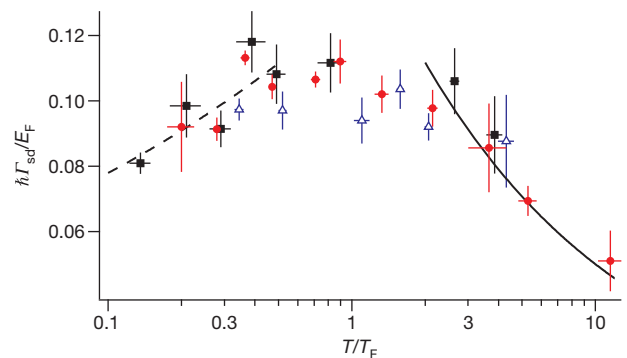

Figure 2 | Spin drag coefficient of a trapped Fermi gas with resonant interactions. The spin drag coefficient Γ_{sd} is normalized by the Fermi rate $E_F/\hbar$ at the trap centre, whereas the temperature is normalized by $T_F = E_F/k_B$. We find agreement between measurements taken at three different axial trapping frequencies, 22.8 Hz (red circles), 37.5 Hz (blue triangles) and 11.2 Hz (black squares). The data for $T/T_F > 2$ fit to a $T^{-1/2}$ law (solid line). Dashed line, a power law fit for $T/T_F < 0.5$ to show the trend. Each point is a mean from typically three determinations of Γ_{sd}, each obtained from a time series of about 30 experimental runs and weighted according to the standard deviation from fitting error and shot to shot fluctuations. Error bars, ± 1 s.e.

$$J_s = -D_s \frac{\partial(n_\uparrow - n_\downarrow)}{\partial_z}$$

where $n_{\uparrow(\downarrow)}$ is the density of spin-up (spin-down) atoms. We calculate J_s using the trap-averaged velocity as $J_s = \frac{1}{2}(n_\uparrow + n_\downarrow)\dot{d}$, where the densities are evaluated at the centre of total mass.

We find that the spin diffusivity is at a minimum when interactions are resonant (see Supplementary Information). The increase in spin diffusivity for positive scattering length a, as well as the decrease in spin drag, argues against the existence of a ferromagnetic state in repulsive Fermi gases, for which diffusion should stop entirely[9,11]. Figure 3 reports the measured spin diffusivity as a function of temperature at unitarity. In the high-temperature limit on resonance, one expects $D_s \propto v/n\sigma \propto T^{3/2}$. At high temperatures, we indeed find this temperature dependence, with a fit giving $D_s = 5.8(2)\frac{\hbar}{m}(T/T_F)^{3/2}$ for $T/T_F > 2$. In the degenerate regime, the spin diffusivity is seen to attain a limiting value of $6.3(3)\hbar/m$.

When comparing these results to theoretical calculations, it is important to account for the inhomogeneous density distributions and velocity profiles. For a homogeneous system on resonance, and at high temperatures compared to the Fermi temperature, we predict $D_s = 1.11\frac{\hbar}{m}(T/T_F)^{3/2}$ and $\Gamma_{sd} = 0.90\frac{E_F}{\hbar}(T/T_F)^{-1/2}$ (see Supplementary Information). The measured spin drag coefficient is smaller by a factor of $0.90/0.16(1) = 5.6(4)$ while the spin diffusivity is larger by about the same factor, $5.8(2)/1.11 = 5.3(2)$, compared to a homogeneous system at the density of the centre of total mass. These factors reflect the inhomogeneity of the system and agree with an estimate from the Boltzmann transport equation (see Supplementary Information). The emergence of a superfluid core at our lowest temperatures will further modify the ratio of trap-averaged to local transport coefficients.

Finally, the measured transport coefficients give for the first time access to the temperature dependence of the spin susceptibility, $\chi_s(T)$, in strongly interacting Fermi gases. Defined as $\chi_s = \frac{\partial(n_\uparrow - n_\downarrow)}{\partial(\mu_\uparrow - \mu_\downarrow)}$, the spin susceptibility describes the spin response to an infinitesimal effective magnetic field or chemical potential difference $\mu_\uparrow - \mu_\downarrow$ applied to the gas, and is a crucial quantity that can discriminate between different states of matter[10]. In a magnetic field gradient, particles with opposite spin are forced apart at a rate determined by the spin conductivity σ_s, while diffusion acts to recombine them. The balance between the processes of diffusion and conduction therefore determines the resulting magnetization gradient, a connection expressed

in the Einstein relation[11] $\chi_s = \sigma_s/D_s$. Assuming the standard relation[11,14] $\sigma_s = n/(m\Gamma_{sd})$,

$$\chi_s = \frac{1}{md\omega_z^2}\frac{\partial(n_\uparrow - n_\downarrow)}{\partial z}$$

where $\frac{\partial(n_\uparrow - n_\downarrow)}{\partial z}$ is evaluated near the trap centre. The inhomogeneous trapping potential does not affect the measurement of χ_s in the hydrodynamic limit at high temperatures (see Supplementary Information). Close to the transition to superfluidity, interaction effects may modify the relation between σ_s and Γ_{sd}.

Figure 4 reports our findings for the spin susceptibility at unitarity, as a function of the dimensionless temperature T/T_F. At high temperatures, we observe the Curie law $\chi_s = n/(k_B T)$, where k_B is Boltzmann's constant. In this classical regime of uncorrelated spins, the susceptibility equals the (normalized) compressibility of the gas $n^2\kappa = \partial n/\partial\mu$ that we also directly obtain from our profiles. At degenerate temperatures, the measured spin susceptibility becomes smaller than the normalized compressibility. This is expected for a Fermi liquid, where $\chi_s = \frac{3n}{2E_F}\frac{1}{1+F_0^a}$ and $\kappa = \frac{3}{2nE_F}\frac{1}{1+F_0^s}$ with Landau parameters F_0^s and F_0^a describing the density (s) and spin (a) response[10]. The spin susceptibility is expected to strongly decrease at sufficiently low temperatures in the superfluid phase, as pairs will form that will not break in the presence of an infinitesimal magnetic field. It is currently debated whether the strongly interacting Fermi gas above the superfluid transition temperature is a Fermi liquid[23] or a state with an excitation gap (pseudogap)[28,29]. The opening of a gap in the excitation spectrum would be revealed as a downturn of the spin susceptibility below a certain temperature. Such a downturn is not observed in χ_s down to $T/T_F \approx 0.2$, and therefore our spin susceptibility data agree down to this point with the expected behaviour for a Fermi liquid.

In conclusion, we have studied spin transport in strongly interacting Fermi gases. The spin diffusivity was found to attain a limiting value of about $6.3\hbar/m$, establishing the quantum limit of diffusion for strongly interacting Fermi gases. Away from resonance, the diffusivity increases.
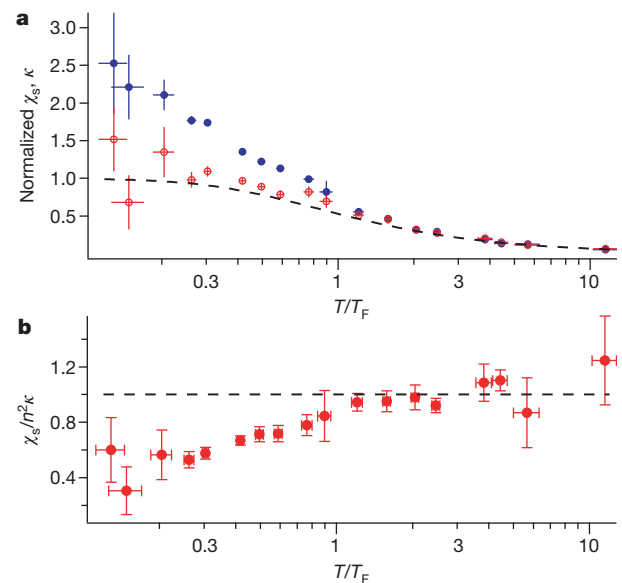

Figure 3 | Spin diffusivity of a trapped Fermi gas. Shown is the spin diffusivity on resonance (D_s, normalized by $\hbar/m$; filled circles) as a function of the dimensionless temperature T/T_F. At high temperatures, D_s obeys the universal $T^{3/2}$ behaviour (solid line). At low temperatures, D_s approaches a constant value of $6.3(3)\hbar/m$ for temperatures below about $0.5T_F$, establishing the quantum limit of spin diffusion for strongly interacting Fermi gases. Error bars, ±1s.e.

Figure 4 | Spin susceptibility on resonance. a, Spin susceptibility (χ_s, open red circles) and isothermal compressibility (κ, filled blue circles), normalized by the values for an ideal Fermi gas at zero temperature. For temperatures below T_F, χ_s becomes suppressed relative to κ, owing to interactions between opposite-spin atoms. Dashed line, χ_s of a non-interacting Fermi gas for comparison. **b,** Red circles, χ_s divided by the value of $n^2\kappa$ obtained from the same clouds. At temperatures above T_F, the ratio of χ_s to $n^2\kappa$ approaches unity (dashed line). Error bars, ±1s.e.

This casts doubt on the possibility of stabilizing a ferromagnetic gas on the repulsive side of the Feshbach resonance[9], which would require a vanishing diffusivity[11]. The observed slow relaxation of spin excitations is a likely explanation for the surprising—possibly non-equilibrium[27]— profiles in spin-imbalanced Fermi gases reported in ref. 30, which did not agree with equilibrium measurements by other workers[20,21,31]. Our measurements of the temperature dependence of the spin susceptibility are consistent with a Fermi liquid picture, and do not reveal a pseudo-gap. An interesting subject for further study is whether spins are still able to diffuse through the superfluid, or whether they travel around it, avoiding the superfluid owing to the pairing gap.

METHODS SUMMARY

The spin mixture is initially prepared at 300 G. To separate the spin components, we reduce the magnetic field to 50 G, where the magnetic moments of the two spin states are unequal, and apply two magnetic field gradient pulses. We then bring the total magnetic field to the Feshbach resonance in about 2 ms.

To reach low temperatures during the approach to equilibrium, evaporative cooling is applied, at 834 G, by gradually lowering the depth of the optical dipole trap. To reach high temperatures, we heat the atoms by switching off the optical dipole trap for up to 3 ms to allow the atom clouds to expand before recapturing them. We then set the final depth of the dipole trap so that the atom number and the temperature remain nearly constant during the approach to equilibrium.

Spin selective imaging is performed by means of *in situ* absorption or phase contrast imaging using two 4-µs imaging pulses separated by 6 µs. These images give the column densities of each spin state, from which we obtain the three-dimensional densities by way of an inverse Abel transform[21]. The gradient in the spin density is obtained from a linear fit to the polarization versus z.

We determine the temperature of the clouds by fitting the density versus potential energy in the vicinity of $z = 0$ (but for all values of the radial coordinate r) to the equation of state of the unitary Fermi gas, measured recently by our group[32]. The trapping potential itself is determined by summing the densities of hundreds of clouds, using the known axial, harmonic trapping potential to convert equidensity lines to equipotential lines and fitting the result to an analytic model.

1. Wolf, S. A. Spintronics: a spin-based electronics vision for the future. *Science* **294**, 1488–1495 (2001).
2. Burrows, A. Neutrinos from supernova explosions. *Annu. Rev. Nucl. Part. Sci.* **40**, 181–212 (1990).
3. Schäfer, T. & Teaney, D. Nearly perfect fluidity: from cold atomic gases to hot quark gluon plasmas. *Rep. Prog. Phys.* **72**, 126001 (2009).
4. Inguscio, M., Ketterle, W. & Salomon, C. (eds) *Ultracold Fermi Gases* (Proc. Int. School of Physics 'Enrico Fermi', Course CLXIV, IOS, 2008).
5. O'Hara, K. M., Hemmer, S. L., Gehm, M. E., Granade, S. R. & Thomas, J. E. Observation of a strongly interacting degenerate Fermi gas of atoms. *Science* **298**, 2179–2182 (2002).
6. Riedl, S. *et al.* Collective oscillations of a Fermi gas in the unitarity limit: temperature effects and the role of pair correlations. *Phys. Rev. A* **78**, 053609 (2008).
7. Cao, C. *et al.* Universal quantum viscosity in a unitary Fermi gas. *Science* **331**, 58–61 (2011).
8. Enss, T., Haussmann, R. & Zwerger, W. Viscosity and scale invariance in the unitary Fermi gas. *Ann. Phys.* **326**, 770–796 (2011).
9. Jo, G.-B. *et al.* Itinerant ferromagnetism in a Fermi gas of ultracold atoms. *Science* **325**, 1521–1524 (2009).
10. Stringari, S. Density and spin response function of a normal Fermi gas at unitarity. *Phys. Rev. Lett.* **102**, 110406 (2009).
11. Duine, R. A., Polini, M., Stoof, H. T. C. & Vignale, G. Spin drag in an ultracold Fermi gas on the verge of ferromagnetic instability. *Phys. Rev. Lett.* **104**, 220403 (2010).
12. D'Amico, I. & Vignale, G. Theory of spin Coulomb drag in spin-polarized transport. *Phys. Rev. B* **62**, 4853–4857 (2000).
13. Weber, C. P. *et al.* Observation of spin Coulomb drag in a two-dimensional electron gas. *Nature* **437**, 1330–1333 (2005).
14. D'Amico, I. & Vignale, G. Coulomb interaction effects in spin-polarized transport. *Phys. Rev. B* **65**, 085109 (2002).
15. Gedik, N., Orenstein, J., Liang, R., Bonn, D. A. & Hardy, W. N. Diffusion of nonequilibrium quasi-particles in a cuprate superconductor. *Science* **300**, 1410–1412 (2003).
16. Garwin, R. L. & Reich, H. A. Self-diffusion and nuclear relaxation in He³. *Phys. Rev.* **115**, 1478–1492 (1959).
17. Anderson, A. C., Edwards, D. O., Roach, W. R., Sarwinski, R. E. & Wheatley, J. C. Thermal and magnetic properties of dilute solutions of He³ in He⁴ at low temperatures. *Phys. Rev. Lett.* **17**, 367–372 (1966).
18. DeMarco, B. & Jin, D. S. Spin excitations in a Fermi gas of atoms. *Phys. Rev. Lett.* **88**, 040405 (2002).
19. Du, X., Luo, L., Clancy, B. & Thomas, J. E. Observation of anomalous spin segregation in a trapped Fermi gas. *Phys. Rev. Lett.* **101**, 150401 (2008).
20. Zwierlein, M. W., Schirotzek, A., Schunck, C. H. & Ketterle, W. Fermionic superfluidity with imbalanced spin populations. *Science* **311**, 492–496 (2006).
21. Shin, Y., Zwierlein, M., Schunck, C., Schirotzek, A. & Ketterle, W. Observation of phase separation in a strongly interacting imbalanced Fermi gas. *Phys. Rev. Lett.* **97**, 030401 (2006).
22. Vichi, L. & Stringari, S. Collective oscillations of an interacting trapped Fermi gas. *Phys. Rev. A* **60**, 4734–4737 (1999).
23. Nascimbène, S., Navon, N., Jiang, K. J., Chevy, F. & Salomon, C. Exploring the thermodynamics of a universal Fermi gas. *Nature* **463**, 1057–1060 (2010).
24. Bruun, G. M., Recati, A., Petchick, C. J., Smith, H. & Stringari, S. Collisional properties of a polarized Fermi gas with resonant interactions. *Phys. Rev. Lett.* **100**, 240406 (2008).
25. Bruun, G. M. Spin diffusion in Fermi gases. *New J. Phys.* **13**, 035005 (2011).
26. Polini, M. & Vignale, G. Spin drag and spin-charge separation in cold Fermi gases. *Phys. Rev. Lett.* **98**, 266403 (2007).
27. Parish, M. M. & Huse, D. A. Evaporative depolarization and spin transport in a unitary trapped Fermi gas. *Phys. Rev. A* **80**, 063605 (2009).
28. Gaebler, J. P. *et al.* Observation of pseudogap behavior in a strongly interacting Fermi gas. *Nature Phys.* **6**, 569–573 (2010).
29. Perali, A. *et al.* Evolution of the normal state of a strongly interacting Fermi gas from a pseudogap phase to a molecular Bose gas. *Phys. Rev. Lett.* **106**, 060402 (2011).
30. Partridge, G. B., Li, W., Kamar, R. I., Liao, Y. & Hulet, R. G. Pairing and phase separation in a polarized Fermi gas. *Science* **311**, 503–505 (2006).
31. Nascimbène, S. *et al.* Collective oscillations of an imbalanced Fermi gas: axial compression modes and polaron effective mass. *Phys. Rev. Lett.* **103**, 170402 (2009).
32. Ku, M. *et al.* Equation of state of a strongly interacting atomic Fermi gas. *Bull. Am. Phys. Soc.* **55**, abstr. W6. 00001 (2010); available at ⟨http://meetings.aps.org/link/BAPS.2010.DAMOP.W6.1⟩ (2010).

Author Contributions All authors contributed to the experimental work. A.S. analysed the data. M.K. developed the algorithm for thermometry. M.W.Z. performed theoretical calculations. A.S. and M.W.Z wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.S. (atsommer@mit.edu).

Transient ferromagnetic–like state mediating ultrafast reversal of antiferromagnetically coupled spins

I. Radu[1,2], K. Vahaplar[1], C. Stamm[2], T. Kachel[2], N. Pontius[2], H. A. Dürr[2,3], T. A. Ostler[4], J. Barker[4], R. F. L. Evans[4], R. W. Chantrell[4], A. Tsukamoto[5,6], A. Itoh[5], A. Kirilyuk[1], Th. Rasing[1] & A. V. Kimel[1]

Ferromagnetic or antiferromagnetic spin ordering is governed by the exchange interaction, the strongest force in magnetism[1–4]. Understanding spin dynamics in magnetic materials is an issue of crucial importance for progress in information processing and recording technology. Usually the dynamics are studied by observing the collective response of exchange-coupled spins, that is, spin resonances, after an external perturbation by a pulse of magnetic field, current or light. The periods of the corresponding resonances range from one nanosecond for ferromagnets down to one picosecond for antiferromagnets. However, virtually nothing is known about the behaviour of spins in a magnetic material after being excited on a timescale faster than that corresponding to the exchange interaction (10–100 fs), that is, in a non-adiabatic way. Here we use the element-specific technique X-ray magnetic circular dichroism to study spin reversal in GdFeCo that is optically excited on a timescale pertinent to the characteristic time of the exchange interaction between Gd and Fe spins. We unexpectedly find that the ultrafast spin reversal in this material, where spins are coupled antiferromagnetically, occurs by way of a transient ferromagnetic-like state. Following the optical excitation, the net magnetizations of the Gd and Fe sublattices rapidly collapse, switch their direction and rebuild their net magnetic moments at substantially different timescales; the net magnetic moment of the Gd sublattice is found to reverse within 1.5 picoseconds, which is substantially slower than the Fe reversal time of 300 femtoseconds. Consequently, a transient state characterized by a temporary parallel alignment of the net Gd and Fe moments emerges, despite their ground-state antiferromagnetic coupling. These surprising observations, supported by atomistic simulations, provide a concept for the possibility of manipulating magnetic order on the timescale of the exchange interaction.

Understanding the physics of non-equilibrium magnetic phenomena is crucial for the understanding of ultrafast magnetization dynamics. This research field started with the seminal observation of subpicosecond demagnetization in ferromagnetic nickel[5] by a 60-fs laser pulse, and was subsequently fuelled by a plethora of fundamentally intriguing ultrafast magnetic phenomena[6]. To study this transient regime of spin dynamics where novel coupling phenomena can emerge, one faces two challenges: (1) how to probe the response of one magnetic moment relative to another; and (2) how to bring the moments out of equilibrium on a timescale corresponding to the exchange interaction.

Ferrimagnetic materials are ideal candidates with which to address both these problems. First, in contrast to elementary ferromagnets, such as Fe, Ni or Co, and typical antiferromagnets like NiO where all the spins are equivalent, ferrimagnets consist of two (or more) non-equivalent and antiferromagnetically coupled spin sublattices[7]. This non-equivalence of the sublattices in combination with an

element-specific detection technique like X-ray magnetic circular dichroism (XMCD) allows us to 'colour' spins in the magnet and to probe the response of one moment relative to another. Second, ultrafast heating of a ferrimagnet over its compensation point in an external magnetic field allows the initiation of the fastest spin reversal reported so far[8,9]. The simplest ferrimagnet, consisting of two collinear magnetic sublattices, may possess a magnetization compensation temperature denoted as T_M. T_M is the result of the different temperature dependences of the two sublattice magnetizations, and represents the point where the magnetizations of these two antiparallel-coupled sublattices are equal and cancel each other. Heating the ferrimagnetic Gd(FeCo) alloy with subpicosecond laser pulses over T_M in an external magnetic field triggers the reversal of the FeCo sublattice within 700 fs (ref. 9). Looking at the Gd sublattice during ultrafast spin reversal of its neighbour Fe would thus be a promising approach to generate and study a state with decoupled behaviour and different dynamics of the two sublattices.

As a model system for such studies, we have used thin films of the amorphous ferrimagnetic alloy $Gd_{25}Fe_{65.6}Co_{9.4}$, where Gd and FeCo represent two antiferromagnetically coupled collinear magnetic sublattices[10]. For simplicity, in the following we focus on the behaviour of the Gd and Fe moments and ignore the small percentage of Co in the transition metal sublattice (Supplementary Information). In order to probe the magnetizations of the two sublattices independently, we used the element-specific XMCD technique. The XMCD measurements were performed in transmission for a fixed X-ray light helicity and opposite orientations of the external magnetic field of ± 0.5 T. The photon energy of the X-ray light was tuned at the Fe L_3 and at the Gd M_5 absorption edges in order to measure their magnetic response separately. In Fig. 1 we show the results of static XMCD measurements as a function of magnetic field measured below and above T_M. The XMCD signals at the Fe and Gd edges reveal the respective sign change of the hysteresis loops on heating the alloy above T_M. Therefore, the data in Fig. 1 confirm that XMCD serves as an element-specific probe of spins in this rare earth-transition metal alloy (see also Supplementary Information).

In order to trigger ultrafast spin dynamics in this alloy, we initiated the reversal of the magnetizations of the two sublattices by ultrafast heating of the sample above T_M using a linearly polarized 60-fs laser pulse with photon energy of 1.55 eV. The dynamics of the Fe and Gd magnetic moments were independently probed using 100-fs soft X-ray pulses[11,12] (see Supplementary Information section 1). Figure 2 shows the results of time-resolved measurements of the dynamics of the Fe and Gd sublattice magnetizations. The measurements were performed with an incident laser fluence of 4.4 mJ cm^{-2} and a sample temperature of 83 K, which is well below $T_M = 250$ K.

First, the net magnetizations of both sublattices rapidly decrease. However, whereas the net magnetization of Fe has collapsed within
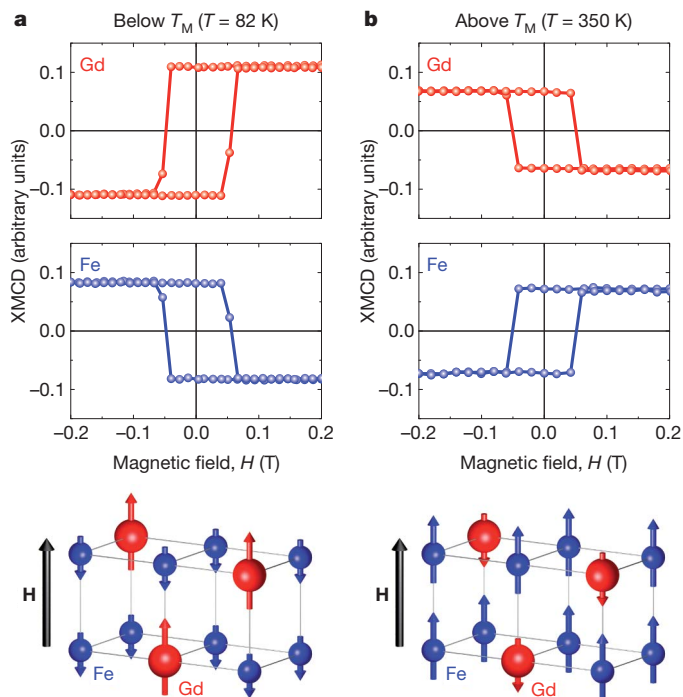
Figure 1 | Ferrimagnetic alignment of the Fe and Gd magnetic moments as measured by element-specific XMCD hysteresis. a, b, Top, XMCD signals measured at the Fe and Gd absorption edges as a function of applied magnetic field below (**a**) and above (**b**) the magnetization compensation temperature (T_M), demonstrating the ferrimagnetic alignment of the Fe and Gd magnetic moments. **a, b,** Bottom, a generic ferrimagnet, showing the alignment of the magnetic moments of the constituent sub-lattices with respect to the external magnetic field, **H**.

300 fs, the demagnetization of Gd takes as long as 1.5 ps. Remarkably, in spite of the exchange coupling between the rare earth and transition metal sublattices, they apparently lose their net magnetizations independently, as would be the case for two decoupled transition metal and rare earth ferromagnets[13,14]. To retrieve the time constants of the involved processes, we have used a double exponential fit function convoluted with the time resolution of the experiment of 100 fs,
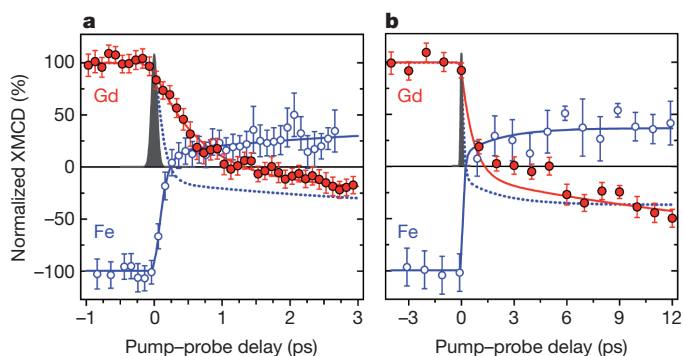
Figure 2 | Element-resolved dynamics of the Fe and Gd magnetic moments measured by time-resolved XMCD with femtosecond time-resolution. a, Transient dynamics of the Fe (open circles) and Gd (filled circles) magnetic moments measured within the first 3 ps. **b,** As **a** but on a 12 ps timescale. Error bars of the experimental data represent the statistical standard error. The measurements were performed at a sample temperature of 83 K for an incident laser fluence of 4.4 mJ cm^{-2}. Experimental time resolution of 100 fs is depicted by the solid Gaussian profile. The solid lines are fits according to a double exponential fit function (Supplementary Information). The dashed line in both panels depicts the magnetization of the Fe sublattice taken with the opposite sign (that is, opposite with respect to the sign of the measured Fe data).

depicted by the Gaussian curve in Fig. 2 (see Supplementary Information section 3). We find a time constant $\tau_{Fe} = 100 \pm 25$ fs for the Fe and $\tau_{Gd} = 430 \pm 100$ fs for the Gd magnetic moment, and thus a ratio $\tau_{Gd}/\tau_{Fe} \approx 4$, characterizing the fast initial drop in the transient dichroic signal. Note that the laser-induced dynamics studied here rely on crossing the compensation temperature T_M and thus are fundamentally different from the laser-induced demagnetization of pure Fe and Gd that rely on crossing the Curie point T_C (refs 15, 16).

Second, the magnetizations of both sublattices switch their directions by crossing the zero signal level and rebuilding their net magnetic moments. Up to 10 ps, the Gd- and Fe-sublattices show distinctly different switching dynamics. Even more surprisingly, within the time-scale between the zero crossings of the Fe and Gd moments (that is, between 300 fs and 1.5 ps), the net Fe and Gd moments are parallel aligned along the z axis despite antiferromagnetic coupling of the spins in the ground state. Note that the net Fe and Gd moments in the transient ferromagnet-like state are reaching rather large values, up to 25% of the equilibrium magnetization. This, together with the substantial laser-induced increase in temperature, indicates a rather strong transient parallel alignment of the Fe and Gd moments. All these observations mean that we have entered a thus far unexplored regime of magnetization dynamics triggered by an ultrashort stimulus, where two exchange-coupled magnetic sublattices are not in equilibrium with each other, show different magnetization dynamics and store different amounts of energy. Consequently, the concept of a magnon (a quantized collective excitation of spins) in such a non-equilibrium ferrimagnetic state should be re-examined, as this state precedes the establishment of the collective Gd-Fe excitations and magnetic resonances in this material.

To further investigate the effect of laser excitation on the degree of switching, we have also studied the dependence of the Fe and Gd dynamics on the laser fluence. As shown in Supplementary Fig. 3, we observe that slightly decreasing the fluence from 4.4 to 4 mJ cm^{-2} leads to longer but still distinct switching times for Fe and Gd. This results in a longer lived transient ferromagnetic-like state, which spans the time range from 700 fs to about 4 ps for this lower fluence.

To obtain a better understanding of the origin of such strongly non-equilibrium spin dynamics, we have developed a model of a ferri-magnet comprising 10^6 localized atomistic exchange-coupled spins and performed numerical simulations of the dynamics after ultra-short laser excitation (Supplementary Information section 4). The model takes into account realistic magnetic moments for Gd and Fe as well as the exchange constants extracted from experimentally measured temperature dependencies of the sublattice magnetizations. In particular, we used an effective Fe–Gd exchange constant of 4.77×10^{-21} J per spin (~140 fs). Following ref. 17, we couple the spin system to the electron temperature, which is calculated using the so-called two-temperature model[18].

Figure 3 shows the results of the simulations when applying a heat pulse with a maximum electron temperature of 1,492 K. First, the time evolution of the sublattice magnetization agrees very well with the experiments, and qualitatively reproduces the timescales for the demagnetization of each sublattice. The time required for the disappearance of the net magnetizations of the sublattices is found to be proportional to the ratio μ/λ, where μ and λ are respectively the magnetic moment of the ion and the damping constant of the sublattice. Second, the ferromagnetic-like state is also reproduced, in a time window close to that observed experimentally. In addition, the reversal of the sublattice magnetizations is found to occur via the mechanism of linear reversal[13], that is, no transverse moment is observed. Note that even making the Gd–Gd exchange interaction in the simulations as strong as the Fe–Fe interaction does not lead to qualitative changes of the dynamics of the Fe and Gd sublattices (Supplementary Information section 5). Moreover, for the sake of argument the model ignores any differences in the electron temperature and the damping constants
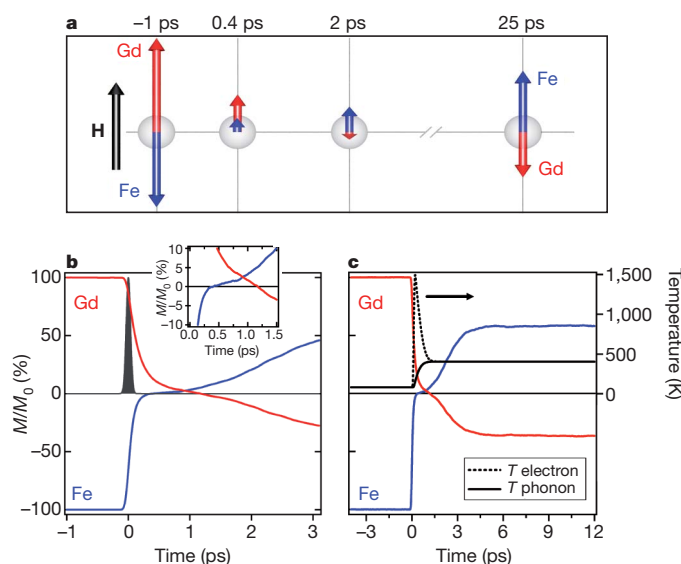
Figure 3 | Computed time-resolved dynamics of the Fe and Gd magnetic moments from localized atomistic spin model. a, Cartoon-like illustration of the non-equilibrium dynamics of the Fe and Gd magnetizations with respect to an external magnetic field **H**. The lengths of the arrows are scaled to the magnitude of the transient XMCD signals shown in Fig. 2. **b, c**, Simulated dynamics for the first 3 ps (**b**) and the first 12 ps (**c**) after laser excitation. The calculations were performed for a peak electronic temperature of 1,492 K with the corresponding transient electronic and phononic temperatures shown in **c**. The transient magnetization changes are normalized to magnetization values at negative delays, that is, to equilibrium values. As is clearly seen, the demagnetization of the Fe is much faster than that of the Gd (see inset in **b**; axes same as main panel). For a time of ~0.5 ps, we observe a parallel alignment of magnetizations of the sublattices. The agreement with the experimental data in Fig. 2 is qualitatively excellent.

for Gd and Fe ($\lambda_{Fe} = \lambda_{Gd} = 0.05$). However, even after such simplifications a good agreement between the experiment and the simulations is observed. This is a strong indication that the novel magnetization dynamics reported here are an intrinsic property of the spins in a ferrimagnet excited by an ultrashort stimulus.

Both the experiments and the simulations show that the ferromagnetic-like state emerges owing to the substantially different dynamics of the transition metal and rare earth sublattices—in particular, the fact that the transition metal reaches the state with zero net magnetization much faster than the rare earth component. The dynamics of the Fe spins in the following time domain occur on the background of a reducing magnetization of the rare earths. The exchange interaction between the Gd and Fe can flip an Fe spin when a Gd spin is reversed. As the exchange is antiferromagnetic, such a process will promote a growth of the net Fe magnetization in the direction parallel to the remaining magnetization of Gd. The simulations indeed confirm that the growth of the Fe magnetization in this time domain occurs with a speed similar (only slightly lower) to that at which the Gd magnetization decreases (Fig. 3b inset).

To conclude, our findings provide unexpected new insights into the fundamentals of ferrimagnetism, showing that two magnetic sublattices may have totally different spin dynamics even on a timescale much longer than the characteristic time of the exchange interaction between them. In the newly discovered transient ferromagnetic-like state, the magnetic system stores a large amount of energy in the intra- and inter-sublattice exchange interactions. Thus it is reasonable to hypothesize that the much weaker interaction of the spins with an external magnetic field hardly influences the spin dynamics in this ultrafast regime, leading to novel switching scenarios. The discovery of these dynamics is important for both the understanding of the physics of ultrafast magnetic phenomena on the timescale of the

exchange interaction, and for the establishment of the fundamental limits on the speed of magnetic recording and information processing.

METHODS SUMMARY

Element-specific hysteresis curves were measured in transmission by setting the X-ray photon energy at the maximum absorption edge (L_3 edge of Fe and M_5 edge of Gd) and sweeping the external magnetic field (which is orientated normal to the sample surface) over ± 0.5 T. The hysteresis was measured at a fixed X-ray light helicity.

Time-resolved XMCD measurements were performed in transmission using 60-fs laser pulses at a wavelength of 800 nm to photo-excite the sample, and using circularly polarized 100-fs X-ray pulses to probe independently the dynamics of the Fe and Gd magnetic moments. The X-ray photon energy was centred at the maximum absorption edge of Fe (L_3) and Gd (M_5) and the laser-induced changes in transmission were measured for opposite orientations of the external magnetic field (± 0.5 T) at a fixed X-ray light helicity. Using a gated detection system, we could record the transient evolution of the pumped (in the presence of laser excitation) and un-pumped (in absence of laser excitation) XMCD signals as a function of pump–probe delay.

For the atomistic spin simulations, the dynamics of the spin system are described using the Landau–Lifshitz–Gilbert (LLG) equation with Langevin dynamics[17]. The model treats GdFeCo as a two-sublattice system with a Heisenberg type of exchange. FeCo is represented by a single generic transition metal sublattice. The exchange constants used in the simulations were obtained by fitting the static, temperature-dependent XMCD measurements for the Fe and Gd sublattices. To simulate the effect of the laser excitation, we couple the spin system to the electron temperature calculated using the two-temperature model[18].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Stöhr, J. & Siegmann, H. C. *Magnetism: From Fundamentals to Nanoscale Dynamics* (Springer, 2006).
2. Weiss, P. L'hypothèse du champ moléculaire et la propriété ferromagnétique. *J. Phys. (Paris)* **6,** 661–689 (1907).
3. Neel, L. Influence de fluctuations du champ moleculaire sur les proprieties magnetiques des corps. *Ann. Phys. (Paris)* **17,** 5–9 (1932).
4. Neel, L. Magnetism and local molecular field. *Science* **174,** 985–992 (1971).
5. Beaurepaire, E., Merle, J.-C., Daunois, A. & Bigot, J.-Y. Ultrafast spin dynamics in ferromagnetic nickel. *Phys. Rev. Lett.* **76,** 4250–4253 (1996).
6. Kirilyuk, A., Kimel, A. V. & Rasing, Th. Ultrafast optical manipulation of magnetic order. *Rev. Mod. Phys.* **82,** 2731–2784 (2010).
7. Gurevich, A. G. & Melkov, G. A. *Magnetization Oscillations and Waves* (CRC, 1996).
8. Aeschlimann, M., Vaterlaus, A., Lutz, M., Stampanoni, M. & Maier, F. Ultrafast thermomagnetic writing processes in rare-earth transition-metal thin films. *J. Appl. Phys.* **67,** 4438–4440 (1990).
9. Stanciu, C. D. *et al.* Subpicosecond magnetization reversal across ferrimagnetic compensation points. *Phys. Rev. Lett.* **99,** 217204 (2007).
10. Stanciu, C. D. *et al.* Ultrafast spin dynamics across compensation points in ferrimagnetic GdFeCo: the role of angular momentum compensation. *Phys. Rev. B* **73,** 220402(R) (2006).
11. Stamm, C. *et al.* Femtosecond modification of electron localization and transfer of angular momentum in nickel. *Nature Mater.* **6,** 740–743 (2007).
12. Khan, S. *et al.* Femtosecond undulator radiation from sliced electron bunches. *Phys. Rev. Lett.* **97,** 074801 (2006).
13. Kazantseva, N. *et al.* Linear and elliptical magnetization reversal close to the Curie temperature. *Europhys. Lett.* **86,** 27006 (2009).
14. Koopmans, B. *et al.* Explaining the paradoxical diversity of ultrafast laser-induced demagnetization. *Nature Mater.* **9,** 259–265 (2010).
15. Bartelt, A. F. *et al.* Element-specific spin and orbital momentum dynamics of Fe/Gd multilayers. *Appl. Phys. Lett.* **90,** 162503 (2007).
16. Wietstruk, M. *et al.* Hot electron driven enhancement of spin-lattice coupling in 4f ferromagnets observed by femtosecond x-ray magnetic circular dichroism. Preprint at ⟨http://arxiv.org/abs/1010.1374⟩ (2010).
17. Kazantseva, N. *et al.* Slow recovery of the magnetization after a sub-picosecond heat pulse. *Europhys. Lett.* **81,** 27004 (2008).
18. Anisimov, S. I., Kapeliovich, B. L. & Perelman, T. L. Electron emission from metal surfaces exposed to ultrashort laser pulses. *Sov. Phys. JETP* **39,** 375–377 (1974).

METHODS

The studied samples are amorphous thin films of $Gd_{25}Fe_{65.6}Co_{9.4}$ of 30 nm thickness which have been deposited by magnetron sputtering on a free-standing Al foil of 500 nm thickness. To avoid oxidation of the GdFeCo layer, Si_3N_4 films of 100 nm and 60 nm thickness were used as buffer and capping layers, respectively. The films exhibit an out-of-plane magnetic anisotropy as deduced from element-specific hysteresis (Fig. 1).

Static and dynamic XMCD measurements have been performed in transmission for opposite orientations of the external magnetic field (± 0.5 T) and at a fixed X-ray light helicity. The magnetic field is oriented perpendicular to the sample and parallel with the X-ray propagation direction. The element-specific hysteresis curves have been recorded by setting the X-ray photon energy at the maximum absorption edge (L_3 edge of Fe and M_5 edge of Gd) and sweeping the external magnetic field between ± 0.5 T.

The time-resolved XMCD measurements have been performed by exciting the sample with 60 fs laser pulses of 800 nm wavelength at a repetition rate of 3 kHz and probing the subsequent dynamics with 100 fs circularly polarized X-ray pulses. By setting the X-ray photon energy at the maximum absorption edge of Fe (L_3 edge) and Gd (M_5 edge) we could disentangle the dynamics of the Fe and Gd magnetic moments. The typical size of the laser beam in focus was 0.8×0.4 mm that led to an incident laser fluence of 4.4 mJ cm^{-2}. The X-ray beam size was 0.3×0.1 mm. The laser-induced changes of X-ray transmission were measured for opposite orientations of the magnetic field using an Si avalanche photodiode and a gated boxcar detection system. The time evolution of the pumped (in the presence of laser excitation) and un-pumped (in absence of laser excitation) signals have been recorded as a function of pump–probe delay. The dynamic XMCD signals have been obtained from the difference of the laser-induced absorption changes measured for opposite magnetic fields. In the main paper we show the ratio of the pumped and un-pumped XMCD curves. The time-resolved XMCD measurements have been performed at a sample temperature of 83 K.

To fit the time-resolved XMCD data, we have used a double-exponential fit function convoluted with the time resolution of the measurement:

$$f(t) = G(t) * [A - B \times (1 - \exp(-t/\tau_1)) - C \times (1 - \exp(-t/\tau_2))]$$

with $G(t)$ describing the Gaussian function that accounts for the time resolution of 100 fs, A representing the value of the transient signal at negative delays, τ_1 and τ_2 the time constants characterizing the two processes—the initial rapid drop and the slower remagnetization in the opposite magnetization direction—describing the temporal evolution of the data, and B and C are exponential amplitudes. Both the static and dynamic XMCD measurements have been performed at the femto-slicing facility at BESSY II Berlin

The dynamics of the spin system are described using the Landau-Lifshitz-Gilbert (LLG) equation with Langevin dynamics, which is a stochastic differential equation with multiplicative noise. The stochastic term augments the local field H_i to incorporate thermal fluctuations. The model uses a numerical solution of the LLG equation:

$$\frac{\partial S_i}{\partial t} = -\frac{\gamma_i}{(1 + \lambda_i^2)\mu_i} \{S_i \times [H_i + \lambda_i S_i \times H_i]\} \tag{1}$$

Here S_i is the spin at the atomic site i, which is normalized to unity, γ_i is the gyromagnetic ratio associated with the spin at site i and μ_i is the magnetic moment associated with each species. H_i is the effective field at each site, which includes contributions, determined by our Hamiltonian, which we choose to be of the generic Heisenberg form, given by:

$$E = -\sum_{\langle ij \rangle} J_{ij} S_i \bullet S_j - \sum_i d_i S_i^z - \sum_i \mu_i B \bullet S_z \tag{2}$$

Here J_{ij} is the exchange constant between neighbouring spins, d_i is the uniaxial anisotropy energy per spin and B is the magnetic field. The field at each site is then given by:

$$H_i = -\frac{\partial E_i}{\partial S_i} \tag{3}$$

which leads to the expression for the fields:

$$H_i = \sum_{nn} S_j + d_i S_i^z + \mu_i B \tag{4}$$

This field must be evaluated for each spin at each step of the numerical integration. We utilize the Heun method for the numerical integration, which is a second order predictor-corrector method.

For the specific problem of our LLG equation for atomistic spins, we augment equation (3) by adding on the thermal term giving the local field in the form:

$$H_i = -\frac{\partial E_i}{\partial S_i} + \varsigma_i \tag{5}$$

with the stochastic term, ς_i, having the properties:

$$\langle \zeta_i^a(t) \zeta_j^b(t) \rangle = 2\delta_{ij}\delta_{ab}(t - t') \frac{\lambda_i k_B T \mu_i}{\gamma_i} \tag{6}$$

and

$$\langle \zeta(t) \rangle = 0 \tag{7}$$

The equation of motion is solved for each spin at each time step. We use an f.c.c. lattice randomly populated by transition metal and rare earth spins, for which we use a uniform random number generator.

The longitudinal relaxation calculations are carried out as follows. Spins are initialized in the same direction and the temperature set to the required value. The evolution of the spin system is calculated by direct numerical integration of the coupled equations of motion. The temporal variation of the magnetization is calculated until equilibrium is reached.

For the switching simulations, the system is subjected to a time varying temperature associated with the conduction electron heat bath. The electron temperature is calculated from the two temperature model[18]:

$$\begin{aligned}
C_e \frac{\partial T_e}{\partial t} &= -\nabla \bullet Q_e - G(T_e - T_1) + S(\mathbf{r}, t) \\
\tau_e \frac{\partial Q_e}{\partial t} + Q_e &= -K_e \nabla T_e \\
C_1 \frac{\partial T_1}{\partial t} &= -\nabla \bullet Q_1 + G(T_e - T_1) \\
\tau_1 \frac{\partial Q_1}{\partial t} + Q_1 &= -K_1 \nabla T_1
\end{aligned} \tag{8}$$

Here C_e is the electron specific heat capacity, C_1 is the lattice/phonon specific heat capacity, and T_e and T_1 are the electron and lattice temperatures, respectively. Q_e and Q_1 are the heat flux vectors through the electron and lattice systems respectively and are (in general) dependent on the position within the body. G is the electron–phonon coupling factor, τ_e and τ_1 are the electron and phonon energy relaxation times, respectively. K_e and K_1 are the thermal conductivity constants of the electron and phonon systems, respectively. We assume uniform heating throughout the sample and a Gaussian laser pulse, which reduces equations (8) to only the first and third equation. This can then be solved using direct numerical integration with a first order Euler scheme.

The values of the physical quantities used in this manuscript are summarized as follows. Fe–Fe exchange, $J_{Fe-Fe} = 2.835 \times 10^{-21}$ J per link. Gd–Gd exchange, $J_{Gd-Gd} = 1.26 \times 10^{-21}$ J per link. Fe–Gd exchange, $J_{Fe-Gd} = -1.09 \times 10^{-21}$ J per link. $\alpha_{Fe} = \alpha_{Gd}$, for which we use 0.05. Though the qualitative result does not change with the reversal, the timescale will vary. Gyromagnetic ratios, $\gamma_{Fe} = \gamma_{Gd} = \gamma$. Magnetic moment of Fe $\mu_{Fe} = 1.92 \mu_B$. Magnetic moment of Gd $\mu_{Gd} = 7.63 \mu_B$. Time step $\Delta t_{real} = 0.1$ fs. Anisotropy constant $d_i = 0.807246 \times 10^{-23}$ J per spin for both the Fe and the Gd.

LETTER

Redox freezing and melting in the Earth's deep mantle resulting from carbon–iron redox coupling

Arno Rohrbach[1] & Max W. Schmidt[1]

Very low seismic velocity anomalies in the Earth's mantle[1,2] may reflect small amounts of melt present in the peridotite matrix, and the onset of melting in the Earth's upper mantle is likely to be triggered by the presence of small amounts of carbonate[3]. Such carbonates stem from subducted oceanic lithosphere in part buried to depths below the 660-kilometre discontinuity and remixed into the mantle. Here we demonstrate that carbonate-induced melting may occur in deeply subducted lithosphere at near-adiabatic temperatures in the Earth's transition zone and lower mantle. We show experimentally that these carbonatite melts are unstable when infiltrating ambient mantle and are reduced to immobile diamond when recycled at depths greater than ~250 kilometres, where mantle redox conditions are determined by the presence of an (Fe,Ni) metal phase[4–6]. This 'redox freezing' process leads to diamond-enriched mantle domains in which the Fe^0, resulting from Fe^{2+} disproportionation in perovskites and garnet, is consumed but the Fe^{3+} preserved. When such carbon-enriched mantle heterogeneities become part of the upwelling mantle, diamond will inevitably react with the Fe^{3+} leading to true carbonatite redox melting at ~660 and ~250 kilometres depth to form deep-seated melts in the Earth's mantle.

The Earth's mantle is a heterogeneous, marble-cake-like composite of pristine, depleted, and mostly pyroxenitic or eclogitic recycled material[7,8], further complicated by heterogeneities left behind in ancient melt passageways (for example, dunites, cumulative pyroxenites, or carbon-enriched domains formed through redox freezing). These heterogeneities are different from average mantle in terms of bulk composition[7], or formed through differences in H_2O or CO_2 volatile influx. Melting processes in the deep mantle are triggered by such heterogeneities, as solidus temperatures for fertile or depleted peridotite are significantly higher than adiabatic temperatures at depths >120 km. Using high pressure experiments, we investigate the formation of C-enriched mantle domains in the vicinity of deeply subducted lithosphere (through redox reactions between recycled carbonate and ambient metal-bearing mantle) and the remelting of such domains when entrained into the convective mantle.

Carbonated peridotite systems are central to understanding processes that involve the recycling of subducted, carbonated lithosphere back into the mantle. Experiments have shown that at pressures above ~2.5 GPa, carbonated mantle melts at lower temperatures than carbonate-free mantle[3,9–12]. However, these experimental studies have maintained carbon in its oxidized form. To understand what happens within the generally reduced deep Earth, it is necessary to include redox equilibria (controlled by oxygen fugacity, f_{O_2}) between intrinsically oxidized carbonate minerals or melts and the reduced and metal-bearing deep mantle.

The redox state of the mantle determines whether carbon is present in its oxidized and potentially mobile form as carbonate or carbonatite melt (which lower the mantle solidus by several hundred degrees[3,9–12]), or whether it is present in its reduced and immobile form as graphite or diamond (which do not affect melting temperatures). In the uppermost mantle to ~250 km depth, f_{O_2} is determined by Fe^{2+}/Fe^{3+} equilibria in silicate minerals. Studies of natural peridotite xenoliths show

that mantle f_{O_2} decreases with increasing pressure[13], such that carbonates or carbonatites are not stable at depths greater than ~120 km in subcratonic and asthenospheric mantle[14,15]. Fe^{2+}/Fe^{3+} equilibria in a predominantly magnesian mantle are inefficient in buffering f_{O_2} because slight changes in the Fe^{2+}/Fe^{3+} ratio have a strong impact on mantle f_{O_2} (ref. 16). Oxygen fugacity in the upper mantle is thus affected by processes such as partial melting, mantle metasomatism, and recycling of oxidized material by subduction. Small amounts of admixed oxidized component may raise the f_{O_2} of a limited mantle domain such that carbonates become stable and f_{O_2} is controlled by equilibria like enstatite + magnesite = olivine + graphite/diamond (EMOG/EMOD)[17]. The onset of carbonatite melting depends then solely on the solidus of carbonated peridotite.

This situation changes at higher pressures: thermodynamic calculations[4] and experiments[5,6,18] demonstrate that f_{O_2} decreases with increasing pressure such that (Fe,Ni)-metal probably becomes stable at ~250 km depth[6,14] and in all mantle regions below[5]. At metal saturation depths, f_{O_2} in the Earth's mantle becomes narrowly constrained. Given equilibrium between mantle phases with molar $Mg/(Mg+Fe^{2+})$ ($=X_{Mg}$) ≈ 0.90 and (Fe,Ni) metal, f_{O_2} can only vary from values around the iron–wüstite (IW) equilibrium where the metal would be Ni-rich, to about 1.5 logarithmic units below IW where the metal would be almost pure iron. Because these equilibria have a considerable buffer capacity and the mantle represents an almost infinite reservoir, the mantle is capable of imposing its ambient f_{O_2} on any additional redox sensitive component, such as carbonates, carbonatites or carbon–hydrogen–oxygen fluids. Redox state differences between such fluids and ambient peridotite may trigger local hydrous redox melting[19] at shallow mantle depths when fluid speciation changes from reduced methane-rich to oxidized water-rich.

At larger mantle depths, small amounts of H_2O can be incorporated into nominally anhydrous minerals[20,21] and CO_2 dominates the lowering of melting temperatures through volatiles[22]. To explore whether redox equilibria involving recycled carbonates trigger the deepest melting in the Earth's mantle, we investigated a carbonated fertile mantle at pressures (10–23 GPa), temperatures (1,400–1,900 °C) and redox conditions relevant for the mid-upper to lower mantle using high-pressure multianvil devices (Methods).

The first set of experiments constrains the solidus of a carbonated fertile peridotite at 10, 14 and 23 GPa to temperatures of 1,535, 1,600 and 1,675 °C (Fig. 1). Melting temperatures determined previously vary considerably, mainly because of variations in the bulk alkali contents and the use of synthetic analogues[3,9–12]. Our 10 GPa solidus temperature corresponds well to the ~1,500 °C reported previously for 10 GPa in a study on natural carbonated peridotite[3], and thus provides a direct link between our data set and lower-pressure data. The solidus from 10 to 23 GPa proceeds parallel to that of alkali-enriched peridotite at 10–20 GPa (ref. 10), although absolute temperatures obtained with our composition are 100–150 °C higher (Fig. 1). This difference results from our considerably lower, peridotite-like, alkali content and a therefore lower alkali/CO_2 ratio (0.06 this study; 0.18 in ref. 10), which is potentially a key variable controlling solidus temperatures[3].

[1]Institut für Geochemie und Petrologie, ETH Zürich, Sonneggstrasse 5, 8092 Zürich, Switzerland.
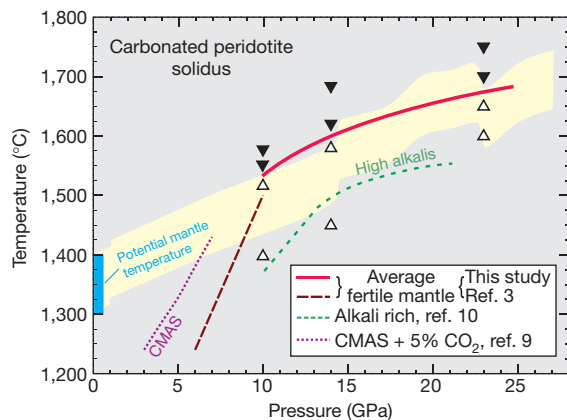
Figure 1 | The solidus of carbonated peridotite. Solid red line, solidus as bracketed by our experiments. Subsolidus runs, open triangles; runs containing carbonatite melt, filled inverted triangles. Run pressure uncertainties are smaller than symbol size; for temperature uncertainties see Methods. Dashed lines, solidi of natural peridotite; dotted line, solidus of a synthetic CaO–MgO–Al$_2$O$_3$–SiO$_2$ (CMAS) + CO$_2$ analogue system[9]. Brown long-dashed line, natural peridotite + 2.5 wt% CO$_2$ (ref. 3); green short-dashed line, natural peridotite + 5 wt% CO$_2$ enriched with 0.5 wt% Na$_2$O + 0.4 wt% K$_2$O (ref. 10). Solidi temperatures decrease with increasing bulk alkali/CO$_2$ ratio[3], although K appears to have a much stronger effect than Na (compare refs 3 and 10 with similar alkali/CO$_2$). Blue bar, range of potential mantle temperatures at 1 bar extrapolated to depth (yellow corridor) assuming adiabatic behaviour in a convecting mantle[32]. The solidus of carbonated peridotite approaches the geotherm at ~10 GPa (ref. 3). As at higher pressures the solidus continues within the range of adiabatic temperatures, we suggest that carbonatitic melts may form at ambient mantle temperatures over a large pressure interval if carbonate is stable with respect to f_{O_2}.

The steep positive slope of the solidus in a pressure–temperature (P–T) plot—55 °C per GPa up to 10 GPa (ref. 3)—changes between 10 and 15 GPa to a rather weak temperature dependence of 12 °C per GPa at >15 GPa. We attribute this change to an increase in activity with pressure of mainly Na$_2$O but also CaO, caused by the continuous decrease in modal abundance of clinopyroxene. Above 15 GPa, pyroxene completely dissolves in majorite–garnet solid solution[23]. Sodium is relatively incompatible in the garnet structure and partitions strongly into a carbonatite liquid, thus causing relatively low melting temperatures. The solidus of carbonated peridotite approaches the mantle geotherm at about 10 GPa and from then on remains close to ambient adiabatic mantle temperatures (Fig. 1). Carbonatite melting in the deep Earth's mantle does therefore not require anomalously high temperatures and carbonatites may be produced over large depth intervals.

In two further series of experiments, we determined carbon speciation as a function of f_{O_2} for slightly subadiabatic mantle temperatures. For this, we first equilibrated the carbonated peridotite with Fe–FeO, Ni–NiO, (Ni,Au)–NiO and Re–ReO$_2$ solid state metal–metal oxide buffers at 14 GPa, 1,450 °C and 23 GPa, 1,600 °C. Fe, Ni, Ni–Au and Re metal were used as capsule materials, and the respective oxide was either contained in the starting material, or (in the case of ReO$_2$) added to the charge. Oxygen fugacities imposed by these buffers range from –1 to +5 log units relative to IW (Methods). Experiments under oxidizing conditions (IW +4 to +5) in Re and Au–Ni capsules yield almost pure magnesite (X_{Mg} = 0.96), whereas runs in Fe and Ni capsules reveal that carbonate is not stable at f_{O_2} < IW +1.2 at either 14 or 23 GPa (Fig. 2). Instead, carbonate is reduced to micrometre-sized diamonds, identified in backscatter electron images and measured using energy-dispersive spectroscopy. Elevated FeO and NiO contents and fractions in the minerals provide further evidence for carbonate reduction via

$$MgCO_3 + 2(Fe,Ni)^0 = 3(Fe,Ni,Mg)O + C \qquad (1)$$

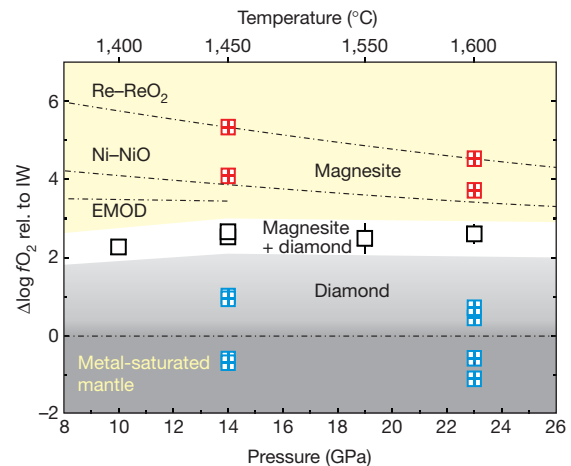where (Fe,Ni,Mg)O is a compound in olivine, garnet and perovskite or is ferropericlase.

Figure 2 | Carbon speciation in natural mantle as a function of pressure, temperature and f_{O_2}. Plotted are calculated log f_{O_2} values relative to the IW reference (Methods) for magnesite-bearing runs (red symbols), diamond-bearing runs (blue symbols) and runs containing diamond + magnesite using Ir as redox sensor (black symbols). Coloured areas mark the stability fields of magnesite (yellow), magnesite+diamond (white) and diamond (light grey). Run durations were at least 24 h to ensure redox equilibrium between the phases (Supplementary Table 1). The stability field of magnesite remains at ≥2 log units above IW up to 23 GPa; no strong pressure dependence is evident from our data. At the low f_{O_2} conditions estimated for the Earth's mantle which lead to metal saturation at pressures higher than ~8 GPa (corresponding to ~250 km depths; dark grey area), carbonates and carbonatites are unstable and will be reduced to diamond according to reaction (1). Relative positions of IW and Ni–NiO buffers calculated after ref. 33; for relative position of Re–ReO$_2$, see Methods. The EMOD buffer dominating in the uppermost mantle[17] (equation given in the text) is limited to ~14 GPa by the stability of olivine and pyroxene. Run pressure uncertainties are within the size of the symbols, as are propagated 2σ errors for f_{O_2} unless error bars are reported. Errors include analytical uncertainties and 2σ standard deviations from averages for runs where f_{O_2} was calculated for multiple phases separately, but do not include uncertainties of the activity models which are difficult to quantify. Because of this, we suggest that an additional uncertainty for all f_{O_2} values reported is ±0.3 log units.

Additionally, we monitored f_{O_2} in experiments saturated in both diamond and carbonate at 10–23 GPa by adding 5 wt% iridium metal as a redox sensor[24]. The Fe content in Ir, together with the FeO content in silicates or oxides, allows the calculation of f_{O_2} conditions during the experiment (Methods). The carbonated peridotite was saturated with elemental carbon through the use of graphite capsules that transformed to diamond or, for noble metal capsules, through addition of 10 wt% C. We also added 3 wt% Fe^0 to enhance the mass reacted through equilibrium (1). These redox sensing experiments yield f_{O_2} values slightly lower than calculated for the EMOD equilibrium at 10 and 14 GPa and then remain rather constant at 2.3–2.7 log units above IW up to 23 GPa. This is consistent with the f_{O_2} corridor between IW +1.2 to +4, as defined by the externally buffered experiments, in which the coexistence of diamond+magnesite would be possible (Fig. 2).

Our results imply that at an ambient, fertile mantle f_{O_2} around IW – 1.5 (ref. 14), carbonate that is remixed into the mantle at >250 km depth is unstable and will be reduced to diamond. Typical carbon concentrations of 20–250 p.p.m. C for sub-ridge mantle (ref. 25 and references therein) would inevitably dissolve in the metal phase present at these depths or form discrete iron carbides, like Fe$_3$C and Fe$_7$C$_3$, depending on Fe–C ratio and P–T conditions[25,26]. Thus, carbonate related melting is unlikely to occur in Earth's lower mantle, the transition zone and the lowermost upper mantle as long as sub-ridge carbon concentrations prevail (Fig. 2). Although the mantle ceases to be metal saturated in shallower parts of the upper mantle (<250 km, ref. 6), the average f_{O_2} up to 100–150 km depth presumably remains too low to sustain carbonates or carbonatites as equilibrium phases at sub-ridge carbon concentrations[15].

However, the addition of subducted carbonate to average mantle fundamentally changes this behaviour. Relative buffer capacity changes in the Earth's mantle induce immobilization of carbonatite melts through 'redox freezing', that is, reduction of carbonatites to diamond, as well as remobilization of carbon through redox melting. Redox melting transforms diamond to carbonatite melts, which potentially control the onset of ultra-deep melting. Starting from a subducting, locally carbonated, relatively oxidized mafic to ultramafic lithosphere, our experiments demonstrate that carbonatite melts will be generated in such lithosphere on thermal relaxation (Fig. 1). This may occur when the lithosphere deflects into the transition zone above the 660-km discontinuity or when stagnating in the lower mantle. On a local scale, oxidized carbonatite melt migrating into the mantle will consume metal to first form iron carbide in an intermittent stage, and then further oxidize the Fe and Ni contained in the carbide to leave a mantle domain that contains all iron as Fe^{2+} and Fe^{3+} in silicates and ferropericlase and all carbon as diamond. Owing to its low viscosity and high wetting properties[27,28], any excess carbonatite not consumed by redox reactions would percolate upwards along grain boundaries and exhaust further (Fe,Ni)-metal and carbide until complete redox freezing—that is, immobilization due to reduction of CO_2 to C^0—is achieved. This presumably very efficient process will eventually exhaust all buffering metal and carbide through precipitation of diamond, and result in a metal-free mantle domain where diamond coexists with Fe^{3+}-bearing garnet or perovskite (Fig. 3).

At the boundary of such domains, where the supply of carbonatite melt does not exceed the redox capacity of Fe,Ni-metal, an iron carbide rim is expected to form. The redox capacity of Fe^{3+} in such mantle domains would be exactly equivalent or slightly superior (due to the Fe^{3+} present before disproportionation of Fe^{2+}) to that necessary to re-convert all diamond to CO_2. Similarly, the maximum increase in C content of mantle domains metasomatized by carbonatites derived from the subducting lithosphere is restricted to ~1,000 p.p.m., equivalent to the amount of carbonatite that may be immobilized by the maximum Fe-metal content of 1 wt% expected in the lower mantle[5].
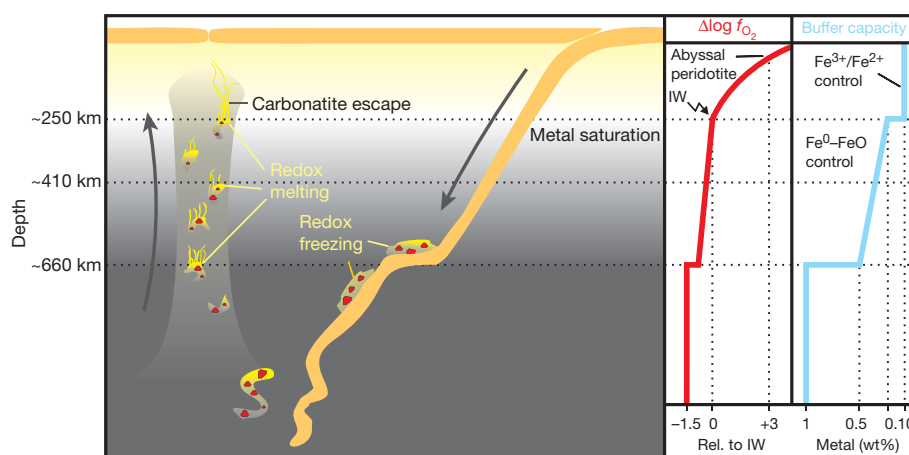
As the metal content is expected to decrease with decreasing pressure (Fig. 3), the corresponding maximum C content of metasomatized domains would also be lower for domains formed at shallower depth.

The inverse process, redox melting, would occur when such mantle heterogeneities are entrained by upwelling mantle and cross the 660-km discontinuity and transition zone. Destabilization of Fe^{3+}-rich perovskite without an adequate amount of Fe^0 for comproportionation (the opposite process to disproportionation) inevitably leads to a sudden increase in Fe^{3+} activity at the 660-km discontinuity and thus to the re-oxidation of diamond to CO_2. This redox reaction would directly result in carbonatite melts within upwelling, previously melt-free but diamond-bearing mantle. Carbonatitic redox melting might also explain the virtual absence of pyrolitic mineral inclusions in diamonds originating from sublithospheric depths (>200–300 km)[29]. Pyrolitic pristine mantle itself does not contain sufficient C to form diamonds[25], and thus diamonds are expected to form dominantly through the above redox freezing process wherever carbonatites percolate. The inverse process—that is, redox melting which destroys most of these diamonds through oxidation—then leads to the generation of melts that carry remnants of such deep mantle domains to the Earth's surface. Average mantle is expected to contain ~1 wt% Fe^0 formed from Fe^{2+} disproportionation in the lower mantle[5]; such metal fractions are stable and would not segregate to Earth's core[30]. For reaction (1), the redox capacity of 1 wt% Fe^0 is equivalent to 0.8 wt% magnesite. As argued above, the properties of carbonatite melt result in a self-regulating mechanism, where infiltrating carbonatite melt oxidizes all Fe^0, leaving behind a diamond-bearing mantle domain with exactly the same redox-capacity (that is, that of 0.8 wt% magnesite). In the reverse process, it can thus be expected that about 1 wt% of carbonatite melt forms in such upwelling mantle domains. Although upwelling occurs at speeds comparable to plate tectonic movements (that is, 1–10 cm yr^{-1}), 1% low viscosity melt in the mantle matrix rises with speeds of at least 10–100 m yr^{-1} (ref. 31). Consequently, carbonatite flow will tend to escape from the upwelling mantle matrix but will suffer redox freezing as long as progressing carbonatite melts encounter

Figure 3 | Carbonatitic redox freezing and redox melting caused by redox capacity changes in Earth's mantle. Main panel, cartoon illustrating a possible sequence of redox freezing and redox melting events driven by oxidation state contrasts between subducted lithosphere and ambient asthenospheric mantle. Right, potential mantle f_{O_2} (red line) and redox buffer capacity (blue line) as function of depth. f_{O_2} decreases in the subcratonic upper mantle[13] to reach ~IW and metal saturation at ~250 km depth[6,14]. Change in redox control from Fe^{2+}/Fe^{3+} in the upper mantle to Fe^0–FeO provides a considerable gain in f_{O_2} buffer capacity. At >250 km depth, the mantle and all redox sensitive components are bound to f_{O_2} levels <IW unless all metal is exhausted. Majoritic garnet in equilibrium with Fe^0 incorporates more Fe^{3+} with increasing pressure[6]; adequate charge balance results in reduction of FeO to Fe^0 which increases the Fe^0 metal abundance. An estimate yields 0.5 wt% metal at the base of the transition zone, assuming all Ni is present as Ni^0 (1,800 p.p.m.) plus ~3,200 p.p.m. Fe^0 calculated from mass balance (mantle: 8 wt% FeO^{total}; 40 vol.% majoritic garnet: 5.5 wt% FeO^{total}, $Fe^{3+}/\Sigma Fe = 0.33$ (ref. 6); 60 vol.% ringwoodite: 9.7 wt% FeO^{total}, $Fe^{3+}/\Sigma Fe = 0.02$ (ref. 14)). Transition to the lower mantle is accompanied by a gain in buffer capacity because MgSi-perovskite incorporates more Fe^{3+} at metal saturation f_{O_2} than majoritic garnet, resulting in ~1.0 wt% metal[5]. Remixing of subducted carbonated lithosphere into the mantle at these depths leads to redox freezing and diamond-enriched mantle domains. Domains incorporated into upwelling mantle will experience redox melting, most prominently at levels where the capacity of the mantle phases to incorporate Fe^{3+} changes drastically, that is, at ~660 km and ~250 km depth. Melts produced from C-enriched domains cannot escape through the surrounding metal-bearing average mantle, but will freeze until redox control shifts from metal saturation back to bulk Fe^{2+}/Fe^{3+} controlled at ~250 m depth.

metal-bearing mantle. It is only at depths $<250\,km$ when garnet becomes less majoritic, Fe^{3+} activity increases, and the mantle ceases to be metal-saturated, that this scenario changes. At this depth, f_{O_2} buffering in the mantle shifts back from being controlled by Fe^0/Fe^{2+} to being controlled by Fe^{2+}/Fe^{3+}, which leads to a dramatic decline in buffering capacity. Such mantle has little potential to hinder carbonatite melts from percolating upwards.

We expect that redox equilibria control the formation of deep carbonatite melts, and that at ~250 km depth a continuous flow of carbonatite melt originating from mantle heterogeneities may form in upwelling mantle, which is probably reflected in seismic low velocity anomalies[1,2] observed at these depths. If melt quantities are too low and evade direct seismological observation, their presence may be confirmed by enhanced electrical conductivity or seismic anisotropies caused by matrix reorganization[3].

METHODS SUMMARY

Starting composition. The starting composition for the experiments represents a fertile peridotite enriched in 5 wt% CO_2. NiO in this composition was set to 1 wt%, to ensure that the NiO content of the phases was high enough for microprobe analysis and f_{O_2} calculations. Experiments using Ir as a redox sensor were without Ni to ensure that the metal is a binary Fe–Ir alloy.

Oxygen fugacity calculations. We calculate f_{O_2} using Fe–FeO and Ni–NiO redox equilibria. f_{O_2} values are reported relative to iron–wüstite (IW), nickel–NiO (NNO) and rhenium–ReO$_2$ (RRO), using activity–composition relationship for (Fe,Ni) alloys and (Au,Ni) alloys (Methods). f_{O_2} values of ferropericlase (fp)-bearing experiments are calculated from

$$2M + O_2 = 2MO \qquad (2)$$

with M = Fe or Ni. f_{O_2} relative to the respective metal–metal oxide (MMO) equilibrium is given by

$$\Delta \log f_{O_2}[MMO] = 2\log a_{MO}^{fp} - 2\log a_M^{metal} \qquad (3)$$

where activity a is defined as molar fraction X times activity coefficient γ. f_{O_2} conditions of runs containing olivine, wadsleyite or ringwoodite were calculated from Fe activity in the olivine polymorph and the coexisting metal phase using binary symmetric solution models for activity corrections.

Iridium redox sensing. Ir metal was used as a redox sensor[24] to determine f_{O_2} in runs containing elemental carbon+magnesite. Margules interaction parameters were obtained by least-squares fitting of a binary asymmetric regular solution model to X-ray data for face-centred cubic (f.c.c.) Ir–Fe alloy (Methods). Activity coefficients γ, dependent on P, T and X for Fe in f.c.c. Ir–Fe alloy, are given by

$$RT\ln(\gamma_{Fe}) = 2X_{Fe}X_{Ir}W_{G\ Ir-Fe} + X_{Ir}^2 W_{G\ Fe-Ir} - 2G^{xs} \qquad (4)$$

with G^{xs} being the excess Gibbs free energy of mixing, and $W_{G\ Ir-Fe}$, $W_{G\ Fe-Ir}$ being the Margules parameters. Correcting the activity model for P and T is described in Methods and shown in Supplementary Fig. 1.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Forsyth, D. W. et al. Imaging the deep seismic structure beneath a mid-ocean ridge: the MELT experiment. Science 280, 1215–1218 (1998).
2. Gu, Y. J., Lerner-Lamb, A. L., Dziewonski, A. M. & Ekström, G. Deep structure and seismic anisotropy beneath the East Pacific Rise. Earth Planet. Sci. Lett. 232, 259–272 (2005).
3. Dasgupta, R. & Hirschmann, M. M. Melting in the Earth's deep upper mantle caused by carbon dioxide. Nature 440, 659–662 (2006).
4. Ballhaus, C. Is the upper mantle metal-saturated? Earth Planet. Sci. Lett. 132, 75–86 (1995).
5. Frost, D. J. et al. Experimental evidence for the existence of iron-rich metal in the Earth's lower mantle. Nature 428, 409–412 (2004).
6. Rohrbach, A. et al. Metal saturation in the upper mantle. Nature 449, 456–458 (2007).
7. Allègre, C. J. & Turcotte, D. L. Implications of a two-component marble-cake mantle. Nature 323, 123–127 (1986).
8. Hofmann, A. W. Mantle geochemistry: the message from oceanic volcanism. Nature 385, 219–229 (1997).
9. Dalton, J. A. & Presnall, D. C. Carbonatitic melts along the solidus of model lherzolite in the system CaO–MgO–Al$_2$O$_3$–SiO$_2$–CO$_2$ from 3 to 7 GPa. Contrib. Mineral. Petrol. 131, 123–135 (1998).
10. Ghosh, S., Ohtani, E., Litasov, K. D. & Terasaki, H. Solidus of carbonated peridotite from 10 to 20 GPa and origin of magnesiocarbonate melt in the Earth's deep mantle. Chem. Geol. 262, 17–28 (2009).
11. Brey, G. P., Bulatov, V. K., Girnis, A. V. & Lahaye, Y. Experimental melting of carbonated peridotite at 6–10 GPa. J. Petrol. 49, 797–821 (2008).
12. Litasov, K. D. & Ohtani, E. Solidus and phase relations of carbonated peridotite in the system CaO–Al$_2$O$_3$–MgO–SiO$_2$–Na$_2$O–CO$_2$ to the lower mantle depths. Phys. Earth Planet. Inter. 177, 46–58 (2009).
13. Woodland, A. B. & Koch, M. Variation in oxygen fugacity with depth in the upper mantle beneath the Kaapvaal craton, South Africa. Earth Planet. Sci. Lett. 214, 295–310 (2003).
14. Frost, D. J. & McCammon, C. A. The redox state of the Earth's mantle. Annu. Rev. Earth Planet. Sci. 36, 389–420 (2008).
15. Stagno, V. & Frost, D. J. Carbon speciation in the asthenosphere: experimental measurements of the redox conditions at which carbonate-bearing melts coexist with graphite or diamond in peridotite assemblages. Earth Planet. Sci. Lett. 300, 72–84 (2010).
16. Gudmundsson, G. & Wood, B. J. Experimental tests of garnet peridotite oxygen barometry. Contrib. Mineral. Petrol. 119, 56–67 (1995).
17. Eggler, D. H. & Baker, D. R. in High Pressure Research in Geophysics (eds Akimoto, S. & Manghnani, M. H.) 237–250 (Centre for Academic Publishing, 1982).
18. O'Neill, H. St C. et al. in Evolution of the Earth and Planets (eds Takahashi, E., Jeanloz, R. & Rubie, D. C.) 73–88 (Geophys. Monogr. 74, Int. Union Geol. Geophys./Am. Geophys. Union, 1993).
19. Taylor, W. R. & Green, D. H. in Magmatic Processes and Physiochemical Principles (ed. Mysen, B. O.) 121–138 (Spec. Publ. No. 1, Geochem. Soc. USA, 1987).
20. Kohn, S. C. Solubility of H$_2$O in nominally anhydrous mantle minerals using ^{1}H MAS NMR. Am. Mineral. 81, 1523–1526 (1996).
21. Keppler, H. & Bolfan–Casanova, N. in Water in Nominally Anhydrous Minerals (eds Keppler, H. & Smyth, J. R.) 193–230 (Reviews in Mineralogy and Geochemistry 62, Min. Soc. Am., 2006).
22. Luth, R. W. in Mantle Petrology: Field Observations and High Pressure Experimentation: A Tribute to Francis R. (Joe) Boyd (eds Fei, Y., Bertka, C. M. & Mysen, B. O.) 297–316 (Geochemical Society, 1999).
23. Ringwood, A. E. The pyroxene–garnet transformation in the Earth's mantle. Earth Planet. Sci. Lett. 2, 255–263 (1967).
24. Woodland, A. B. & O'Neill, H. St C. Thermodynamic data for Fe-bearing phases obtained using noble metal alloys as redox sensors. Geochim. Cosmochim. Acta 61, 4359–4366 (1997).
25. Dasgupta, R. & Hirschmann, M. M. The deep carbon cycle and melting in Earth's interior. Earth Planet. Sci. Lett. 298, 1–13 (2010).
26. Lord, O. T., Walter, M. J., Dasgupta, R., Walker, D. & Clark, S. M. Melting in the Fe–C system to 70 GPa. Earth Planet. Sci. Lett. 284, 157–167 (2009).
27. Minarik, W. G. & Watson, E. B. Interconnectivity of carbonate melt at low melt fraction. Earth Planet. Sci. Lett. 133, 423–437 (1995).
28. Hunter, R. H. & McKenzie, D. The equilibrium geometry of carbonate melts in rocks of mantle composition. Earth Planet. Sci. Lett. 92, 347–356 (1989).
29. Stachel, T., Brey, G. P. & Harris, J. W. Inclusions in sublithospheric diamonds: glimpses of deep earth. Elements 1, 73–78 (2005).
30. Bagdassarov, N., Solferino, G., Golabek, G. J. & Schmidt, M. W. Centrifuge assisted percolation of Fe–S melts in partially molten peridotite: time constraints for planetary core formation. Earth Planet. Sci. Lett. 288, 84–95 (2009).
31. Connolly, J. A. D., Schmidt, M. W., Solferino, G. & Bagdassarov, N. Permeability of asthenospheric mantle and melt extraction rates at mid-ocean ridges. Nature 462, 209–212 (2009).
32. Stixrude, L. & Lithgow-Bertelloni, C. Influence of phase transformations on lateral heterogeneity and dynamics in Earth's mantle. Earth Planet. Sci. Lett. 263, 45–55 (2007).
33. Campbell, A. J. et al. High pressure effects on the iron–iron oxide and nickel–nickel oxide oxygen fugacity buffers. Earth Planet. Sci. Lett. 286, 556–564 (2009).

METHODS

Starting material. The starting compositions for multianvil experiments used in this study represent a fertile peridotite composition[34] enriched in 5 wt% CO_2 (Supplementary Table 2). Starting composition pyr5 was used to determine the solidus temperature of carbonated peridotite and for runs where the oxygen fugacity of the charge was controlled by the Fe, Ni, and (Au,Ni) capsule materials. Both NiO and CoO contents of pyr5 were set to 1 wt%, to ensure that NiO contents of the phases were high enough for microprobe analysis and f_{O_2} calculation from these data. Experiments using Ir metal as redox sensor were performed with the simplified pyrx5 starting material to ensure that the metal phase in the charge is a binary Fe–Ir alloy.

Starting composition pyr5 was prepared from reagent grade oxides SiO_2, TiO_2, Al_2O_3, Cr_2O_3, MgO, fired overnight at 1,000 °C before mixing and grinding under acetone in an agate mortar. Part of the CaO inventory was added to the oxide mix as reagent grade $CaCO_3$ (dried at 220 °C for 5 h), mixed and decarbonized by stepwise heating to 1,000 °C. Afterwards, iron was added as synthetic fayalite, NiO and CoO as oxides, all dried at 220 °C for 5 h. CO_2 was added as a mixture of reagent grade $CaCO_3$, $MgCO_3$ and Na_2CO_3 dried at 220 °C for 5 h in 61/34/5 wt ratio. All components were mixed and ground again thoroughly under acetone in an agate mortar.

Starting composition pyrx5 was prepared from reagent grade SiO_2, Al_2O_3, MgO (all fired at 1,000 °C for 5 h) synthetic fayalite, synthetic wollastonite, natural magnesite (kindly provided by P. Ulmer), and reagent grade Na_2CO_3 (all dried at 220 °C for 5 h).

To ensure nominally anhydrous conditions, the starting materials were stored in an desiccator together with silica gel; synthetic rock powders were dried for 2 h at 220 °C before pressing it into capsules and the filled capsule was dried again for 1 h at 220 °C before final welding.

Multianvil experiments. High pressure experiments (Supplementary Table 1) were performed in 600 t and 1,000 t Walker type multianvil devices. We used 32-mm tungsten carbide cubes as second stage anvils and natural pyrophyllite gaskets as pressure transmitting media. We used an 18/11 assembly for runs at 10 GPa (18 mm octahedron edge length, 11 mm truncated edge length), a 14/8 assembly for 14 GPa and a 10/3.5 assembly for 19 and 23 GPa. All assemblies consist of a Cr_2O_3 doped MgO octahedron with $LaCrO_3$ heater (stepped in case of 18/11 and 14/8), ZrO_2 thermal insulation, MgO spacers, and Mo metal plugs to ensure electrical contact between assembly and WC cubes. Pressure was computer controlled during the entire run; temperature was measured with B- or C-type thermocouples and controlled with a Eurotherm controller to about ±5 °C. Experiments were quenched (by turning off the power supply) at a rate of ~800 °C per second. The total temperature precision including temperature gradients within the capsule might be ±15 °C in 18/11 and 14/8 experiments. Temperature gradients within the 10/3.5 assembly might be 30–50 °C over the capsule length of ~1.2 mm. Details of pressure calibration, design, and temperature gradients of the 10/3.5 assembly are given in ref. 35.

Encapsulation. Runs to determine the carbonated peridotite solidus were performed in welded $Au_{80}Pd_{20}$ capsules. No attempt was made to control f_{O_2} during these runs. The use of conventional Pt–C double capsules appeared to be problematic as the space is very limited and recovering the sample from the graphite capsule that transformed to diamond during the runs was rather challenging. Mass balance calculations and measured Fe contents in the capsule material show that iron loss from the charge to the $Au_{80}Pd_{20}$ capsule is negligible (see also X_{Mg} values in Supplementary Table 2).

Runs using Ir metal as redox sensor were performed in welded $Au_{80}Pd_{20}$ capsules. We added ~10 wt% of graphite to the charges to obtain the f_{O_2} buffering through coexisting C^0 and carbonate in a peridotite matrix. We added ~5 wt% fine grained Ir metal and 3–5 wt% of metallic Fe to ensure redox equilibrium via carbonate reduction according to reaction (1).

Slightly lower X_{Mg} of all phases in f_{O_2} monitoring runs (Supplementary Table 2) are indicative for reaction (1) and therefore for the attainment of redox equilibrium in these runs. Both diamond and magnesite were present in experimental charges in direct contact with each other as verified by backscattered imaging, EDX and EPMA analysis. This encapsulation method yields results that agree within error to those from experiments in C capsules surrounded by Re metal foil (compare f_{O_2} of runs 39 and 47; Supplementary Table 5) but sample recovery and polishing was facilitated considerably, especially for runs at 23 GPa.

Fe^0 and Ni^0 capsules were not welded but cold sealed at run pressure with a conical cap. We added ~3 wt% of the respective metal to the charge to facilitate attainment of equilibrium.

(Au,Ni) alloy capsules were prepared as welded Au outer capsule and Ni foil inner capsule which formed an homogeneous (Au,Ni) alloy at run pressures and temperatures. In runs 22 and 29b we added ~10 wt% NiO in order to raise the NiO content of the phases and accordingly relative f_{O_2}. This NiO-enriched compositions crystallized a NiO-rich periclase instead of ferropericlase at 23 GPa (run 29b) and an olivine enriched in Ni_2SiO_4-component at 14 GPa (run 22, see Supplementary Table 2).

Analytical methods. Recovered sample capsules (in case of 10/3.5 runs, complete octahedra) were mounted in an epoxy resin, ground down with sand paper and polished with polycrystalline diamond paste. Runs to determine the solidus were polished under dry conditions using synthetic polishing paper.

Runs were analysed with a Jeol JXA8200 microprobe using minerals and synthetic oxides as standards. Results are given in Supplementary Table 2. Silicates and oxides were measured at 15 kV and 20 nA with a focused beam, counting times were 20 s on the peak and 10 s on the background. Perovskites and carbonates were measured with an electron beam of 5 μm diameter at 10 nA to minimize radiation damage. Note, however that CaSi-perovskite was still unstable under the beam, which results in low totals. Additionally, CaSi-perovskite often shows a blurred exsolution texture that is difficult to analyse without obtaining mix signals. The majority of the experiments, especially under subsolidus conditions (despite run durations up to 4 days) showed average grain sizes between 3 and 10 μm and did not allow a broad beam. In experiments AR13 and AR18, some phases were too small to be measured by electron microprobe and were identified using backscattered electron images and EDX (Jeol JSM 6390LA scanning electron microscope).

Metals were measured with a focused beam at 15 kV and 20 nA, counting times were 30 s on the peak and 10 s on the background. We used Fe, Ir, Ni and Co metal as standards for quantification to reduce matrix effects. To minimize the excitation volume of the sample, to be able to measure small Ir-Fe alloy grains, we measured Ir at 15 kV and 20 nA using its low excitation energy Mα peak.

Calculation of oxygen fugacities. We calculated f_{O_2} of the experimental charges using Fe–FeO and Ni–NiO redox equilibria. f_{O_2} values are reported relative to iron–wüstite (IW) and nickel–NiO (NNO) buffers. The relative positions of IW and NNO at P and T were calculated after ref. 33, the relative position of the Re–ReO$_2$ (RRO) buffer was obtained from ref. 36. The activity–composition relationship for (Fe,Ni) alloys were taken from ref. 37, activity data for (Au,Ni) alloys were adopted from ref. 38.

Periclase-bearing experiments. Oxygen fugacities of ferropericlase (fp)-bearing experiments (19 and 23 GPa) were calculated using equilibria of the form

$$2M + O_2 = 2MO \tag{5}$$

with M = Fe or Ni. f_{O_2} relative to the respective metal–metal oxide (MMO) equilibrium reads:

$$\Delta \log f_{O_2} \,[\text{MMO}] = 2 \log a_{MO}^{fp} - 2 \log a_M^{metal} \tag{6}$$

where activity (a) is defined as molar fraction (X) times activity coefficient (γ). The molar fraction for metals and phases, calculated activity coefficients and calculated f_{O_2} values of the experiments are summarized in Supplementary Tables 3–5. a_{FeO}^{fp} was determined using a binary regular solution model:

$$RT\ln(\gamma_{FeO}^{fp}) = (11{,}000 + 0.011P)(1 - X_{FeO}^{fp})^2 \tag{7}$$

for FeO, where P is in bar, T in K, and R is the gas constant[39]. We note that ferric iron content of ferropericlase was not measured and microprobe FeO^{total} data were used to calculate X_{Fe}. The $Fe^{3+}/\Sigma Fe$ for ferropericlase in systems with $X_{Mg} = 0.9$ however is generally <0.05 (refs 40–42), so calculated f_{O_2} values might therefore be overestimated by a maximum of 0.1 log units. a_{NiO}^{fp} was calculated using a ternary regular solution model for NiO:

$$RT\ln(\gamma_{NiO}^{fp}) = W_{Ni\text{-}Mg}(X_{Mg})^2 + W_{Ni\text{-}Mg}(X_{Fe})^2 \\ + (W_{Ni\text{-}Mg} + W_{Ni\text{-}Fe} - W_{Mg\text{-}Fe})X_{Mg} \tag{8}$$

where T is in K, and R is the gas constant[43,44].

Olivine-, wadsleyite- or ringwoodite-bearing experiments. f_{O_2} conditions relative to the quartz–fayalite–iron (QFI) equilibrium of runs containing olivine, wadsleyite or ringwoodite (10, 14 and 19 GPa) were calculated using the following expressions:

$$2Fe^0 + O_2 + SiO_2 = Fe_2SiO_4 \tag{9}$$

$$\Delta \log f_{O_2} \,[\text{QFI}] = 2 \log a_{Fe}^{ol/wad/rw} - 2 \log a_{Fe}^{metal} - \log a_{SiO_2} \tag{10}$$

Activity coefficients were calculated using a binary symmetric solution model where:

$$RT\ln(\gamma_{Fe}^{ol}) = W_{Fe\text{-}Mg}^{ol}(1 - X_{Fe}^{ol})^2 \tag{11}$$

Margules parameters including W_V to correct the olivine interaction parameter for pressure were taken from ref. 45 (Supplementary Table 6).

NiO-bearing olivines show a slight negative deviation from ideal mixing, even at the high temperatures relevant for our experiments[46,47]. Owing to this we set $\gamma_{NiO}^{ol} = 0.9$; but the effect on calculated f_{O_2} is minor, that is, in the range ± 0.05 log units. Silica activities (a_{SiO_2}) in experiments containing olivine and clinoenstatite were calculated from the equilibrium:

$$Mg_2SiO_4 + SiO_2 = Mg_2Si_2O_6 \qquad (12)$$

where

$$\log a_{SiO_2} = \log a_{en}^{opx} - \log a_{fo}^{ol} - \log K \qquad (13)$$

Enstatite activity in pyroxene was calculated after ref. 48 and forsterite activity in olivine after ref. 45 as described above for fayalite. $\log K$ of reaction (12) was calculated using thermodynamic data[49]. f_{O_2} values relative to IW are obtained by adding the difference in f_{O_2} between QFI and IW equilibria at run P and T to $\Delta \log f_{O_2}$ [QFI].

Activity–composition relation in the system Fe-Ir at high P and T. Iridium metal was used as a redox sensor[24] to monitor f_{O_2} in runs that contain both elemental carbon and magnesite between 10 and 23 GPa. Margules interaction parameters (W) dependent on P, T and X are calculated according to the expression

$$W_G = W_{H\ 1bar} - TW_S + (P - 1)W_V \qquad (14)$$

where T is in K and P in bar[50].

$W_{H\ 1bar}$ was obtained by least-squares fitting of a binary asymmetric regular solution model[51] to the non-temperature-dependent part of the excess Gibbs energy (G^{xs}) for face centred cubic (f.c.c.) Ir–Fe alloy[52]. W_S is $-5\,J\,mol^{-1}\,K^{-1}$ according to the temperature dependence of the excess Gibbs energy expression for f.c.c. Ir–Fe alloy[52]. W_V is estimated by least-squares fitting of a binary asymmetric regular solution model to excess volumes of the Fe-Ir alloy derived from X-ray data[52]. Activity coefficients γ dependent on P and T, and X values for Fe in f.c.c. Ir–Fe alloy result from a binary asymmetric regular solution model[51]:

$$RT\ln(\gamma_{Fe}^{metal}) = 2X_{Fe}X_{Ir}W_{G\ Ir-Fe} + (X_{Ir})^2 W_{G\ Fe-Ir} - 2G^{xs} \qquad (15)$$

where

$$G^{xs} = X_{Fe}X_{Ir}(X_{Ir}W_{G\ Fe-Ir} + X_{Fe}W_{G\ Ir-Fe}) \qquad (16)$$

The effect of correcting the activity model for P and T is shown in Supplementary Fig. 1.

34. Palme, H. & O'Neill, H. St C. in *The Mantle and The Core* Vol 2. (ed. Carlson, R. W.) 1–38 (Elsevier-Pergamon, 2003).

35. Stewart, A. J., van Westrenen, W., Schmidt, M. W. & Melekhova, E. Effect of gasketing and assembly design: a novel 10/3.5 mm multi-anvil assembly reaching perovskite pressures. *High Press. Res.* **26,** 293–299 (2006).

36. Campbell, A. J. *et al.* Pressure–volume–temperature studies of metal–oxide pairs. In *COMPRES Annual Meeting* (2007); ⟨www.geol.umd.edu/~ajc/Posters/CampbellCOMPRES2007poster.pdf⟩.

37. Predel, B. Fe–Ni in *The Landolt–Börnstein Database* (ed. Madelung, O.) (Springer Materials, 1995); doi:10.1007/10474837_1321.

38. Wang, J., Lu, X.-G., Sundman, B. & Su, X. Thermodynamic assessment of the Au–Ni system. *Comput. Coupling Phase Diagr. Thermochem.* **29,** 263–268 (2005).

39. Frost, D. J. Fe^{2+}–Mg partitioning between garnet, magnesiowüstite and $(Mg,Fe)_2SiO_4$ phases of the transition zone. *Am. Mineral.* **88,** 387–397 (2003).

40. McCammon, C. A., Peyronneau, J. & Poirier, J.-P. Low ferric iron content of $(Mg,Fe)O$ at high pressures and temperatures. *Geophys. Res. Lett.* **25,** 1589–1592 (1998).

41. Frost, D. J., Langenhorst, F. & van Aken, P. Fe–Mg partitioning between ringwoodite and magnesiowüstite and the effect of pressure, temperature and oxygen fugacity. *Phys. Chem. Miner.* **28,** 455–470 (2001).

42. Frost, D. J. & Langenhorst, F. The effect of Al_2O_3 on Fe–Mg partitioning between magnesiowüstite and magnesium silicate perovskite. *Earth Planet. Sci. Lett.* **199,** 227–241 (2002).

43. Seifert, S. & O'Neill, H. St C. Experimental determination of activity–composition relations in Ni_2SiO_4–Mg_2SiO_4 and Co_2SiO_4–Mg_2SiO_4 olivine solid solutions at 1200 K and 0.1 MPa and 1573 K and 0.5 GPa. *Geochim. Cosmochim. Acta* **51,** 97–104 (1987).

44. O'Neill, H., St. C., Canil, D. & Rubie, D. C. Oxide metal equilibria to 2500 °C and 25 GPa: implications for core formation and the light component in the Earth's core. *J. Geophys. Res.* **103,** 12239–12260 (1998).

45. Frost, D. J. The structure and sharpness of $(Mg,Fe)_2SiO_4$ phase transformations in the transition zone. *Earth Planet. Sci. Lett.* **216,** 313–328 (2003b).

46. Hirschmann, M. Thermodynamics of multicomponent olivines and the solution properties of $(Ni,Mg,Fe)_2SiO_4$ and $(Ca,Mg,Fe)_2SiO_4$ olivines. *Am. Mineral.* **76,** 1232–1248 (1991).

47. Hirschmann, M. M. & Ghiorso, M. S. Activities of nickel, cobalt and manganese silicate in magmatic liquids and application to olivine/liquid and to silicate/melt partitioning. *Geochim. Cosmochim. Acta* **58,** 4109–4126 (1994).

48. O'Neill, H. S. C. & Wall, V. J. The olivine-orthopyroxene-spinel oxygen geobarometer, the nickel precipitation curve, and the oxygen fugacity of the Earth's upper mantle. *J. Petrol.* **28,** 1169–1191 (1987).

49. Barin, I., Sauert, F., Schultze-Rhonhof, E. & Sheng, W. S. *Thermochemical Data of Pure Substances* Parts I and II (Weinheim, 1989).

50. Wood, B. J. & Fraser, D. G. *Elementary Thermodynamics for Geologists* (Oxford Univ. Press, 1976).

51. Mukhopadhyay, B., Basu, S. & Holdaway, M. J. A discussion of Margules-type formulations for multicomponent solutions with a generalized approach. *Geochim. Cosmochim. Acta* **57,** 277–283 (1993).

52. Swartzendruber, L. J. The Fe–Ir (iron–iridium) system. *Bull. Alloy Phase Diagrams* **5,** 48–52 (1984).

LETTER

Distinct representations of olfactory information in different cortical centres

Dara L. Sosulski[1], Maria Lissitsyna Bloom[1]†, Tyler Cutforth[1]†, Richard Axel[1] & Sandeep Robert Datta[1]†

Sensory information is transmitted to the brain where it must be processed to translate stimulus features into appropriate behavioural output. In the olfactory system, distributed neural activity in the nose is converted into a segregated map in the olfactory bulb[1–3]. Here we investigate how this ordered representation is transformed in higher olfactory centres in mice. We have developed a tracing strategy to define the neural circuits that convey information from individual glomeruli in the olfactory bulb to the piriform cortex and the cortical amygdala. The spatial order in the bulb is discarded in the piriform cortex; axons from individual glomeruli project diffusely to the piriform without apparent spatial preference. In the cortical amygdala, we observe broad patches of projections that are spatially stereotyped for individual glomeruli. These projections to the amygdala are overlapping and afford the opportunity for spatially localized integration of information from multiple glomeruli. The identification of a distributive pattern of projections to the piriform and stereotyped projections to the amygdala provides an anatomical context for the generation of learned and innate behaviours.

In vision and touch, information central to perception is ordered in space in the external world and this order is maintained from the peripheral sense organs to the cortex. Olfactory information, however, does not exhibit a discernible spatial order in the physical world and this poses the question of how odours are represented in the brain. In mammals, olfactory perception is initiated by the recognition of odorant molecules by a large repertoire of receptors in the olfactory sensory epithelium[4]. Individual olfactory sensory neurons express one of approximately 1,000 receptors[5–7], and each receptor interacts with multiple odorants. Neurons expressing a given receptor, although randomly distributed within zones of the olfactory epithelium, project with precision to two spatially invariant glomeruli in the olfactory bulb[1–3]. Thus, the randomly distributed population of neurons activated by an odorant in the olfactory epithelium is consolidated into a discrete stereotyped map of glomerular activity in the olfactory bulb[8,9].

This highly ordered map of spatially invariant glomeruli must then be transformed in higher olfactory centres. The projection neurons of the olfactory bulb, mitral and tufted cells, extend an apical dendrite into a single glomerulus and send axons to several telencephalic areas, including a significant input to the piriform cortex and cortical amygdala[10–15]. Electrophysiological studies and optical imaging reveal that individual odorants activate subpopulations of neurons distributed across the piriform without spatial preference[16,17]. The piriform therefore discards the spatial segregation of the bulb and returns to a highly dispersed organization in which different odorants activate unique ensembles of cortical neurons. However, the patterns of neural activity do not allow us to discern whether mitral and tufted cell projections from a given glomerulus to cortical neurons are segregated or distributed, and whether they are random or determined. Distinguishing between these possibilities is important for understanding odour perception because a random representation of odour identity could accommodate learned olfactory

behaviours, but is unlikely to specify innate behaviours. Rather, innate olfactory behaviours are likely to result from the activation of genetically determined, stereotyped neural circuits. We have therefore developed a strategy to trace the projections from identified glomeruli in the olfactory bulb to higher olfactory cortical centres.

Mitral and tufted cells that innervate a single glomerulus were labelled by electroporation of tetramethylrhodamine (TMR)-dextran under the guidance of a two-photon microscope. This technique labels mitral and tufted cells that innervate a single glomerulus and is sufficiently robust to allow the identification of axon termini within multiple higher order olfactory centres (Figs 1a–c, 2 and Supplementary Figs 1–4). Labelling of glomeruli in the olfactory bulbs of mice that express GFP under the control of specific odorant receptor promoters permits us to examine potential stereotypy of projections from identical glomeruli (MOR28–IRES–GFP ($n = 8$), MOR1-3–IRES–GFP ($n = 13$), and MOR174-9–IRES–GFP ($n = 10$))[18]. Labelling of random unidentified glomeruli was performed in mice expressing synapto-pHluorin (OMP–IRES–spH), a fluorescent marker that allows us to visualize individual glomeruli[9]. This permits us to sample the projection patterns from multiple different glomeruli (Fig. 1c).

Electroporation of a single glomerulus results in the labelling of 6–17 neurons in the mitral cell layer (mean = 9.2 ± 0.8; Fig. 1e and
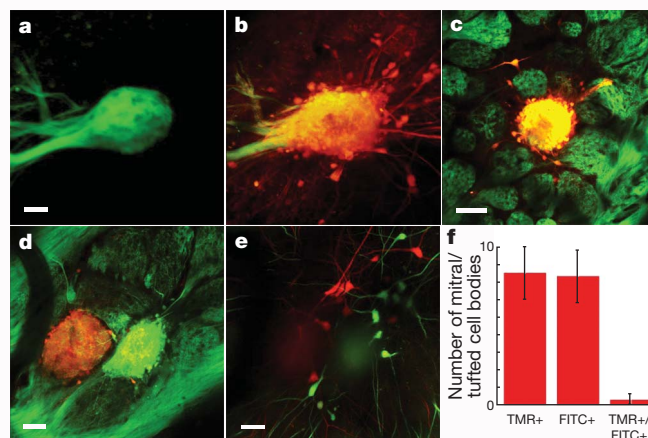

Figure 1 | Targeted electroporation of TMR-dextran labels cells that innervate a single glomerulus in the olfactory bulb. a, b, A mouse olfactory bulb in which the MOR174-9 glomerulus is labelled with GFP, before (**a**) and after (**b**) electroporation with TMR-dextran. Scale bar, 40 μm. **c**, Image similar to **a** where electroporation was performed in a mouse in which spH is expressed in all glomeruli (OMP–IRES–spH, green); note that labelling (red) is confined to a single glomerulus. Scale bar, 85 μm. **d**, Control experiment in an OMP–IRES–spH mouse in which neighbouring glomeruli were electroporated with TMR-dextran (red, left) and FITC-dextran (green, right). Scale bar, 45 μm. **e**, Labelling of mitral cells (red, green) as a result of the experiment in **d**. **f**, Quantification of the overlap in mitral cell labelling in experiments similar to **d** (error bars represent s.e.m.; $n = 4$).

[1]Department of Neuroscience and the Howard Hughes Medical Institute, College of Physicians and Surgeons, Columbia University, New York, New York 10032, USA. †Present addresses: Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, USA (M.L.B., S.R.D.); Department of Molecular, Cell and Developmental Biology, University of California, Santa Cruz, California 95064, USA (T.C.).
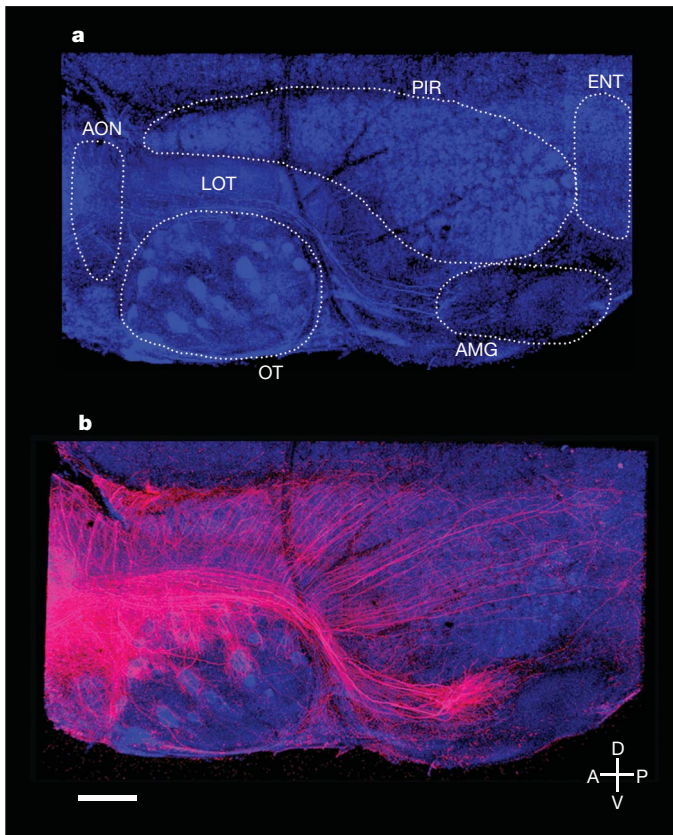
Figure 2 | Mitral/tufted cells connected to a single glomerulus show distinct patterns of projections to several areas of the olfactory cortex. a, A flattened hemi-brain preparation of the olfactory cortex with nuclei identified by counterstain (blue, NeuroTrace 435) and relevant structures outlined in white. AON, anterior olfactory nucleus; AMG, cortical amygdala; ENT, lateral entorhinal cortex; LOT, lateral olfactory tract; OT, olfactory tubercle; PIR, piriform cortex. **b**, A hemi-brain from a mouse in which a single glomerulus was electroporated with TMR-dextran (red). Note the unique pattern of projection in each of the olfactory areas. A, anterior; P, posterior; D, dorsal; V, ventral. Scale bar, 700 μm. See also Supplementary Fig. 3.

Supplementary Figs 1c, 2; see Methods). We have demonstrated that all mitral and tufted cells labelled in this manner innervate a single glomerulus by electroporating TMR-dextran into one glomerulus and fluorescein isothiocyanate (FITC)-conjugated dextran into a neighbouring glomerulus (Fig. 1d). Examination of the mitral and tufted cells after this two-colour electroporation reveals either red or green mitral and tufted cells with only a rare cell labelled simultaneously with the two dyes (1/60 cells, $n = 4$) (Fig. 1e, f). This labelling strategy therefore restricts incorporation of tracer to mitral and tufted cells innervating only one glomerulus.

We observe that projections from individual glomeruli extend to all major olfactory cortical regions including the accessory olfactory nucleus, piriform cortex, olfactory tubercle, cortical amygdala and lateral entorhinal cortex (Fig. 2b and Supplementary Fig. 3). Visualization of the extent of axonal projections was facilitated by the development of a flattened hemi-brain preparation that enables high-resolution imaging of all olfactory centres except the anterior olfactory nucleus, which is obscured by the overlying lateral olfactory tract (LOT). The glomeruli we examined project to all the major olfactory cortical regions independent of the spatial location of the glomerulus within the olfactory bulb ($n = 21$ different glomeruli in the flattened preparation; see later). Each of the different higher olfactory centres receives a qualitatively unique pattern of input from the olfactory bulb (Fig. 2b and Supplementary Fig. 3). In the piriform cortex a distributive representation is observed, whereas in the amygdala mitral cell projections are broad but spatially segregated.

Mitral and tufted cell axons extend to the piriform cortex via the LOT. We observe that axonal branches exit the LOT at right angles and extend upward to densely and diffusely project to the piriform cortex along the entire anteroposterior axis (Fig. 3a–c), with no apparent spatial preference in any dimension. High-resolution multiphoton imaging reveals varicosities likely to be axonal boutons (Supplementary Fig. 4). The spatial distribution of these varicosities is similar in every field imaged and is independent of glomerular origin, indicating that mitral and tufted cell synapses with piriform neurons are distributed throughout the piriform cortex (Supplementary Fig. 5). The density of these varicosities within the piriform cortex is also similar regardless of the identity of the electroporated glomerulus, further suggesting that each glomerulus makes a similar number of synapses (MOR1-3, $10.2 \pm 0.57\,\mu$m of axon per varicosity; M72 mouse 1, 9.9 ± 0.65; and M72 mouse 2, $10.1 \pm 0.36\,\mu$m of axon per varicosity; Supplementary Figs 4, 5). The highly dispersed pattern of projection to the piriform cortex is observed from every glomerulus examined independent of its identity or location within the olfactory bulb ($n = 21$; Supplementary Fig. 6). On visual inspection, the patterns of projection from two identical glomeruli are no more similar than the patterns of projection observed from two different glomeruli.

We performed hierarchical and k-means clustering to determine whether the observed patterns of projections from different glomeruli are quantitatively distinguishable (see Methods). We were unable to identify any parameters, including axon fibre positions, density of TMR labelling, centre of mass X and Y coordinates and centroid X and Y coordinates, that reliably distinguish the projection patterns from different glomeruli (Supplementary Table 1). All measured parameters were similar on comparison of the projection patterns from identical or different glomeruli (Supplementary Fig. 7).

Cross-correlation analysis was performed to compare the patterns of piriform projections from different glomeruli (see Methods and Supplementary Fig. 8 for detailed explanation of method and interpretation). The correlograms comparing identical and distinct glomeruli show an extended region of moderate correlation (Fig. 3d–f). These data indicate that the dispersed pattern of projections is largely homogeneous in density over several spatial scales (Supplementary Fig. 9) and indicate that the patterns are similar for each of the 24 glomeruli we examined. The similarity of correlograms from identical and different glomeruli provides further evidence that the pattern of piriform projections does not differ for each of the distinct glomerulus types. Thus, the mitral cells innervating an individual glomerulus discard the insular and invariant spatial segregation of the bulb and project dense, dispersed axons to the piriform cortex with no discernible spatial bias.

Next we examined the patterns of projections of single glomeruli to the cortical amygdala. The cortical amygdala consists of three nuclei: the anterior cortical, posterolateral cortical and posteromedial cortical nuclei. The posteromedial cortical nucleus, a major site of innervation from the accessory olfactory bulb[19], receives no discernible input from any of the glomeruli of the main olfactory bulb we examined (Fig. 4a–c, d–f, $n = 33$ glomeruli in the en face preparation, see Methods), and we observe relatively sparse projections anterior to the posterolateral nucleus. We therefore restrict our analysis to the posterolateral cortical nucleus. The patterns of projection from individual glomeruli in the posterolateral nucleus reveal dense, patchy axonal projections that have a focal nexus surrounded by a less dense halo of fibres (Fig. 4a–f). Despite the diffuse nature of projections, different glomeruli appear to send fibres to anatomically distinct and spatially invariant regions of the posterolateral cortical amygdala (Fig. 4a–f). For example, projections from the MOR1-3 glomerulus consistently occupy the most medial aspect of the posterolateral nucleus, whereas projections from the MOR28 glomerulus terminate more laterally. These conclusions are evident on visual inspection and are supported by more quantitative analysis.

k-means clustering using relevant parameters extracted from the amygdala projection patterns after image alignment (centre of mass

a MOR1-3 **b** MOR174-9 **c** Random spH

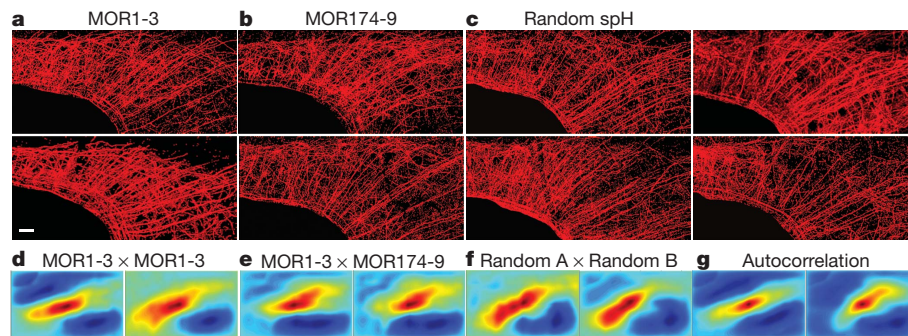**d** MOR1-3 × MOR1-3 **e** MOR1-3 × MOR174-9 **f** Random A × Random B **g** Autocorrelation

Figure 3 | Projections from single glomeruli to piriform cortex are disperse, homogeneous and indistinguishable. a–c, Images of axons innervating the piriform cortex (red) from mitral and tufted cells that connect to the glomerulus corresponding to MOR1-3 (**a**), MOR174-9 (**b**) or a random selection of glomeruli labelled with TMR-dextran (**c**). Scale bar, 500 μm. **d–f**, Correlograms plotted using the matrix of correlation coefficients generated by normalized cross-correlation of two MOR1-3 piriforms (**d**), a MOR1-3 and a MOR174-9 piriform (**e**), and two piriforms in which random glomeruli were labelled (**f**). Cross-correlation is performed using aligned images of projection patterns as seen in **a–c**. **g**, Autocorrelograms generated using methods from **d** in which a labelled piriform is compared to itself. Note that correlograms in **g** are essentially indistinguishable from the correlograms in **d–f**.

X coordinate, X position of medial-most fibre, absolute medial fibre density, ratio of lateral/medial fibre density) was performed on the projections from three identified glomeruli, MOR1-3 ($n = 5$), MOR174-9 ($n = 5$) and MOR28 ($n = 4$). This cluster analysis correctly assigns glomerular identity for 79% of the samples examined (one-way MANOVA, $P = 0.0006$, $\alpha = 0.05$).

We performed normalized cross-correlation analysis to compare further the projection patterns from different glomeruli. Cross-correlation analysis reveals a single peak in the correlogram that reflects the more focal nature of projections to this brain region than in the piriform (Fig. 4g–j). The correlograms between the projection patterns of identical glomeruli reveal peaks that show a small displacement from the centre (Fig. 4g, h, Supplementary Figs 10, 11a–c, 12g, h and Supplementary Tables 2, 3). Cross-correlation using images of projection patterns from different glomeruli show more varied and often very large displacements (Fig. 4i, Supplementary Figs 10, 11d–j, 12i and Supplementary Tables 2, 3). These data indicate that the cortical amygdala receives spatially stereotyped projections from individual glomeruli. Although individual glomeruli project to fixed positions, extensive overlap is observed for the projections from different glomeruli. The apparently random pattern of projections in the piriform and the determined pattern in the amygdala are likely to

provide the anatomical substrates for distinct olfactory-driven behaviours mediated by these two brain regions.

Insight into the logic of olfactory perception will depend on an understanding of how the highly ordered glomerular map is represented in higher olfactory centres. Previous experiments have used the injection of tracer molecules into the bulb or cortex to relate the spatial position of projection neurons in the bulb with their targets in higher olfactory centres[10–15]. These experiments predate the ability to identify specific glomeruli[2], precluding a determination as to whether projections from a single mitral/tufted cell or a single glomerulus are random or stereotyped.

We have defined a neural circuit that conveys olfactory information from specific glomeruli in the olfactory bulb to the piriform cortex and the cortical amygdala. A distributive representation of neurons in the sensory epithelium is converted into a topographical map in the bulb on the convergence of like axons onto spatially invariant glomeruli[1,2]. The piriform discards this spatial order; axons from individual glomeruli project diffusely to the piriform without apparent spatial preference. Neurons from every glomerulus elaborate similar axonal arborizations and quantitative analyses fail to identify features that may distinguish the individual projection patterns. These data are in accordance with retrograde tracings using rabies virus that reveal the convergence of multiple, spatially distributed glomeruli on a small number of piriform
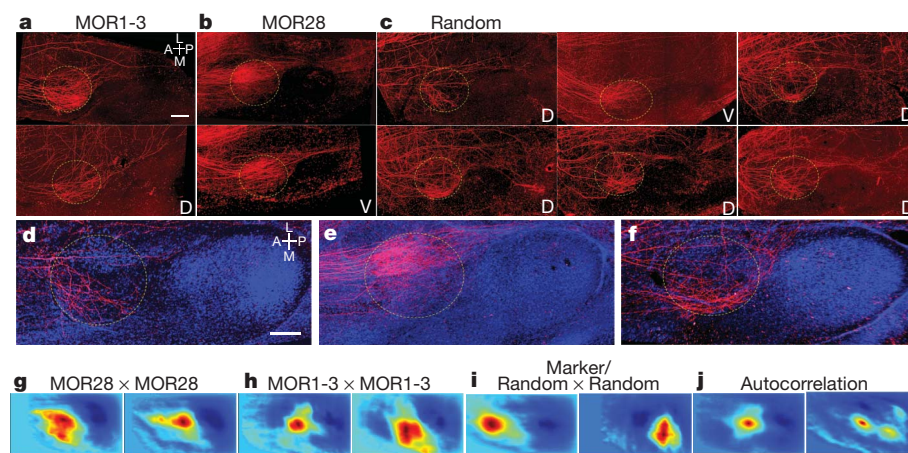

a MOR1-3 **b** MOR28 **c** Random

d **e** **f**

g MOR28 × MOR28 **h** MOR1-3 × MOR1-3 **i** Marker/Random × Random **j** Autocorrelation

Figure 4 | Projections from single glomeruli to the cortical amygdala are broad, patchy and stereotyped. a–c, Images of the cortical amygdala reveal similar projections from the mitral and tufted cells that connect to the MOR1-3 glomerulus in two different brains (circle represents the approximate posterolateral cortical nucleus boundary) (**a**), but projections that are distinct from those of mitral/tufted cells connected to the MOR28 glomerulus (**b**) or six randomly selected glomeruli (**c**). 'D' or 'V' in the bottom right corner of the image indicates whether the electroporated glomerulus was located dorsally or ventrally in the bulb. M, medial; L, lateral. Scale bar, 400 μm. **d–f**, Counterstained images

from a subregion of images in **a–c** showing a closer view of projection patterns. Scale bar, 400 μm. **g–i**, Correlograms plotted using the matrix of correlation coefficients generated by normalized cross-correlation of MOR28 × MOR28 projection patterns within the posterolateral cortical amygdala (PLCo) (**g**), MOR1-3 × MOR1-3 projection patterns (**h**), or projection patterns from glomeruli of different types (**i**). **j**, Autocorrelograms of the PLCo from two labelled glomeruli correlated with themselves. Note that in the en bloc preparation shown here, the lateral/medial axis (indicated by the orientation bars) is synonymous with the dorsal/ventral axis, as this region of brain is curved.

neurons[20]. A distributive representation of glomerular projections in the piriform is also observed on anterograde tracing of individual mitral cells after Sindbis virus infection[21].

Optical imaging and electrophysiological studies of neural responses to odours reflect these anatomical transformations. Distributed neural activity in the sensory epithelium of the nose is transformed in the bulb, with each odour eliciting distinct spatial patterns of glomerular activity[8,9]. A second transformation is apparent in the piriform cortex where individual odorants activate unique ensembles of neurons that are distributed without discernible spatial order[16,17]. These neurons also exhibit discontinuous receptive fields; neurons within an ensemble responsive to a given odour respond to multiple, structurally dissimilar odours[16,22]. The dispersed projections to the piriform provide an anatomical substrate for the generation of these patterns of neural activity.

One model consistent with both the anatomy and physiology invokes the random convergence of excitatory inputs from mitral cells onto piriform neurons such that each piriform neuron would sample a random combination of glomerular inputs. If the connections from bulb to cortex are indeed random, then the representation of the quality of an odorant or its valence in the piriform must be imposed by experience. Odorants, however, can elicit innate behavioural responses, indicating that a second area of the brain must receive determined inputs from the olfactory bulb. The pattern of projections to the posterolateral amygdala implicates this structure in the generation of innate olfactory-driven behaviours. This suggestion is in accordance with the finding that disruption of the amygdala abrogates innate, odour-driven behaviours but leaves learned olfactory responses intact[23,24].

The olfactory circuits we describe in the mouse are reminiscent of the architecture of the olfactory system in *Drosophila*[25], despite the six hundred million years of evolution that separate the two organisms. In *Drosophila*, neurons expressing a given odorant receptor are distributed throughout the antenna and converge on spatially invariant glomeruli in the antennal lobe. Information from the antennal lobe bifurcates, with one branch exhibiting spatially invariant projections to the lateral horn, a brain region mediating innate olfactory behaviours. A second branch projects to the mushroom body, a structure required for learned olfactory responses. This anatomical and functional bifurcation provides a context in which to consider the generation of the various forms of olfactory-driven behaviour in both flies and mice. Our data indicate that innate olfactory behaviours derive from determined neural circuits selected over evolutionary time, whereas learned behaviours may be mediated by the selection and reinforcement of random ensembles of neurons over the life of an organism.

METHODS SUMMARY

Single glomeruli and their associated mitral and tufted cells were labelled in adult mice via electroporation of 3 kD TMR-dextran (Invitrogen) under a two-photon microscope (Prairie Technologies). Glass electrodes were backfilled with dye and guided to a glomerulus centre, and current was applied to the back of the pipette (50 V, 30-ms pulses, 2 Hz, 7–10 min, repeated 2–4 times, Grass Technologies). Afterwards, mice were allowed to recover for 5 days before being killed. Excised brains were prepared for imaging either by flattening the cortex of the right hemisphere or surgical excision of the cortical amygdala. TMR-dextran signal was amplified by antibody staining (Jackson Immunoresearch), and samples were counterstained (NeuroTrace 435, Invitrogen). Images were acquired using multiphoton excitation on a Zeiss 710 or a Prairie *in vivo* microscope. Image z-stacks of the piriform cortex or cortical amygdala were aligned using SPM8 and Photoshop, and then quantified using ImageJ and custom-written Matlab code (Mathworks).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Ressler, K. J., Sullivan, S. L. & Buck, L. B. Information coding in the olfactory system: evidence for a stereotyped and highly organized epitope map in the olfactory bulb. *Cell* **79**, 1245–1255 (1994).
2. Mombaerts, P. *et al.* Visualizing an olfactory sensory map. *Cell* **87**, 675–686 (1996).
3. Vassar, R. *et al.* Topographic organization of sensory projections to the olfactory bulb. *Cell* **79**, 981–991 (1994).
4. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
5. Malnic, B., Hirono, J., Sato, T. & Buck, L. B. Combinatorial receptor codes for odors. *Cell* **96**, 713–723 (1999).
6. Chess, A., Simon, I., Cedar, H. & Axel, R. Allelic inactivation regulates olfactory receptor gene expression. *Cell* **78**, 823–834 (1994).
7. Niimura, Y. & Nei, M. Evolutionary changes of the number of olfactory receptor genes in the human and mouse lineages. *Gene* **346**, 23–28 (2005).
8. Rubin, B. D. & Katz, L. C. Optical imaging of odorant representations in the mammalian olfactory bulb. *Neuron* **23**, 499–511 (1999).
9. Bozza, T., McGann, J. P., Mombaerts, P. & Wachowiak, M. *In vivo* imaging of neuronal activity by targeted expression of a genetically encoded probe in the mouse. *Neuron* **42**, 9–21 (2004).
10. Haberly, L. B. & Price, J. L. The axonal projection patterns of the mitral and tufted cells of the olfactory bulb in the rat. *Brain Res.* **129**, 152–157 (1977).
11. Scott, J. W., McBride, R. L. & Schneider, S. P. The organization of projections from the olfactory bulb to the piriform cortex and olfactory tubercle in the rat. *J. Comp. Neurol.* **194**, 519–534 (1980).
12. Price, J. L. An autoradiographic study of complementary laminar patterns of termination of afferent fibers to the olfactory cortex. *J. Comp. Neurol.* **150**, 87–108 (1973).
13. Luskin, M. B. & Price, J. L. The distribution of axon collaterals from the olfactory bulb and the nucleus of the horizontal limb of the diagonal band to the olfactory cortex, demonstrated by double retrograde labeling techniques. *J. Comp. Neurol.* **209**, 249–263 (1982).
14. Buonviso, N., Revial, M. F. & Jourdan, F. The projections of mitral cells from small local regions of the olfactory bulb: an anterograde tracing study using PHA-L (*Phaseolus vulgaris* Leucoagglutinin). *Eur. J. Neurosci.* **3**, 493–500 (1991).
15. Ojima, H., Mori, K. & Kishi, K. The trajectory of mitral cell axons in the rabbit olfactory cortex revealed by intracellular HRP injection. *J. Comp. Neurol.* **230**, 77–87 (1984).
16. Stettler, D. D. & Axel, R. Representations of odor in the piriform cortex. *Neuron* **63**, 854–864 (2009).
17. Rennaker, R. L., Chen, C. F., Ruyle, A. M., Sloan, A. M. & Wilson, D. A. Spatial and temporal distribution of odorant-evoked activity in the piriform cortex. *J. Neurosci.* **27**, 1534–1542 (2007).
18. Shykind, B. M. *et al.* Gene switching and the stability of odorant receptor gene choice. *Cell* **117**, 801–815 (2004).
19. de Olmos, J., Hardy, H. & Heimer, L. The afferent connections of the main and the accessory olfactory bulb formations in the rat: an experimental HRP-study. *J. Comp. Neurol.* **181**, 213–244 (1978).
20. Miyamichi, K. *et al.* Cortical representations of olfactory input by trans-synaptic tracings. *Nature* advance online publication, doi:10.1038/nature09714 (22 December 2010).
21. Ghosh, S. *et al.* Sensory maps in the olfactory cortex defined by long-range viral tracing of single neurons. *Nature* advance online publication doi:10.1038/nature09945 (30 March 2011).
22. Poo, C. & Isaacson, J. S. Odor representations in olfactory cortex: ''sparse'' coding, global inhibition, and oscillations. *Neuron* **62**, 850–861 (2009).
23. Blanchard, D. C. & Blanchard, R. J. Innate and conditioned reactions to threat in rats with amygdaloid lesions. *J. Comp. Physiol. Psychol.* **81**, 281–290 (1972).
24. Slotnick, B. M. Olfactory discrimination in rats with anterior amygdala lesions. *Behav. Neurosci.* **99**, 956–963 (1985).
25. Vosshall, L. B. & Stocker, R. F. Molecular architecture of smell and taste in *Drosophila*. *Annu. Rev. Neurosci.* **30**, 505–533 (2007).

Author Contributions S.R.D., D.L.S. and R.A conceived the project, participated in its development and wrote the manuscript. S.R.D. and D.L.S. developed methods and performed all experiments and data analysis. T.C. generated the MOR1-3 and MOR174-9–IRES–GFP mice. M.L.B. performed mouse husbandry and immunostaining.

METHODS

Surgery. Mice were anaesthetized with ketamine/xylazine (100 mg kg^{-1} or 10 mg kg^{-1}, respectively; Sigma-Aldrich), and temperature was maintained at 37 °C on a feedback-controlled heating pad (Fine Science Tools). The scalp was removed, and membrane overlying the skull was cleared using a microblade (Roboz). An aluminium headpost cut from square bar (Small Parts) was attached to the skull using RelyX luting cement (Henry Schein). The borders of the exposure were covered with silicone sealant (VWR). For dorsal glomeruli, the skull overlying the olfactory bulb was thinned using a dental drill (KaVo) and removed with forceps, and the dura was peeled back using fine forceps (Roboz). For lateral glomeruli, the skin overlying the cheek and zygomatic bone was removed, and vessels were cauterized (Fine Science Tools). The muscle attached to the zygomatic was peeled away, and the bone was removed with microscissors (Roboz). The eye and surrounding tissue was removed with microscissors; bleeding was stopped using gelfoam (Henry Schein), and mice were administered 0.7 ml Ringer's solution (Henry Schein). The skull overlying the bulb was thinned and removed, and the dura peeled away. After electroporation, the bulb was coverslipped and covered in 2% agarose (Sigma-Aldrich), the exposure was covered in lidocaine jelly (Henry Schein) and then silicone sealant. Buprenorphine (0.05 mg kg^{-1}, Henry Schein) was administered after the mouse could right itself. Mice recovered for 5 days after electroporation, and were then deeply anaesthetized with ketamine/xylazine and killed by paraformaldehyde perfusion.

Electroporation. Mice were placed under a two-photon microscope (Ultima, Prairie Technologies), and a ×16 objective was used to focus on a single glomerulus (.8NA, Nikon). A Ti-Sapphire laser (Coherent) was tuned to 880 nm for experiments. Pulled glass pipets (Sutter, 5–6 μm tip) were backfilled with either lysine-fixable TMR-dextran (3 kDa, 12.5 mg ml^{-1} in PBS) with biotin or lysine-fixable FITC-dextran (3,000 MW, 12.5 mg ml^{-1} in PBS) with biotin (Invitrogen), and filled halfway with 0.9% w/v NaCl. The pipette was mounted on an electrode holder (WPI)/manipulator (Luigs and Neumann), and its tip was directed to the three-dimensional centre of a glomerulus under two-photon guidance. Current was applied to the pipette using a stimulator (50 V, 30-ms pulses, at 2 Hz, repeated 2–4 times, Grass SD-9 stimulator). The black lead of the stimulator was connected to the animal via an alligator clip on the foot. Note that although we observed robust and reliable long-range diffusion/transport of TMR-dextran from labelled mitral and tufted cells to the olfactory cortex under these electroporation conditions, FITC-dextran showed substantively less diffusion/transport (for example, few or no labelled fibres in the LOT), which precluded the use of FITC-dextran to explore axonal projections from the bulb to the cortex. Mice in which the glomeruli were not clearly labelled, the labelling was non-specific (an exceedingly rare occurance) or in which labelled mitral cells in the mitral cell layer could not be identified during the course of the experiment were excluded from subsequent analysis.

Histological processing. Mice were killed by transcardial perfusion with 13 ml PBS, followed by 10 ml 1% paraformaldehyde. Brains were extracted and processed differentially depending on the region to be imaged. For imaging of the whole olfactory cortex, the left hemisphere was discarded and the subcortical matter of the right hemisphere was removed using forceps. Cortical tissue above the rhinal sulcus was dissected away, and guide cuts were made with a needle blade (Fine Science Tools). The ventral hemisphere (which includes all of the structures in the olfactory cortex) was then flattened between two slides separated by a 600-μm spacer constructed of #1.5 coverslips. The sample was placed in 4% paraformaldehyde overnight, and then stored in PBS until antibody staining. For imaging of the amygdala, the brain was dissected as previously described, and a needle blade was used to cut along the LOT to excise the olfactory tubercle; an additional cut was made above the cortical nuclei of the amygdala to excise the olfactory amygdala. We use this en bloc preparation of the amygdala to avoid distortions that occur at the ventral edge of the flattened hemi-preparation, where the cortical amygdala resides. This distortion is minimal in the centre of the flattened preparation where the piriform cortex is located. The amygdala was fixed in 4% paraformaldehyde overnight, and then stored in PBS until antibody staining. Before application of antibody, samples were soaked in glycine (1.87 g per 500 ml PBS, Sigma-Aldrich) for 2 h, followed by 1 h in PBS, 30 min in sodium borohydride (0.4 g per 400 ml PBS, Sigma-Aldrich) and 1 h in PBS to quench autofluorescence. The following series of antibodies and washes were then used: primary antibody for 3 days (1:500 rabbit anti-TMR, Invitrogen) in block (2% Triton X-100, 1% bovine serum albumin in PBS), wash 1 day (2% Triton X-100/PBS), secondary antibody for 2 days (1:250 goat anti-rabbit in block, Jackson Immunoresearch), wash 1 day, and tertiary antibody (1:250 TMR-conjugated donkey anti-goat in block, Jackson Immunoresearch) and counterstain (1:150 NeuroTrace 435 in block, Invitrogen) for 2 days, followed by 1 day of wash. Samples were mounted in Vectashield (Vector Labs) on a slide with the shallow end of a Lab-Tek chambered coverglass as a coverslip (NUNC).

Quantification of cellular labelling. Two different methods were used to quantify the efficiency of our labelling method. First, after electroporation, we acquired z-stacks from the surface of the olfactory bulb through the mitral cell layer of the bulb (see Supplementary Movie 1), and used these z-stacks to count the number of cell bodies labelled in the mitral cell layer of the olfactory bulb post hoc. We were unable to quantify the number of tufted cells labelled for technical reasons: the brightness and number of cells labelled superior to the mitral cell layer, as well as the brightness of the glomerulus, made it difficult to quantify cell bodies (Supplementary Fig. 1), and the diversity in morphology and location of tufted cells made it difficult to identify them based on visual and depth criteria. Second, previous work has indicated that mitral cell axons travel superior to tufted cell axons in a segregated manner in the LOT and, unlike tufted cell axons, project to areas posterior to the anterior piriform cortex and olfactory tubercle[26–30]. We observed that near the most posterior aspect of the olfactory tubercle, the axons of presumed mitral cells form a distinct bundle that continues to travel past the tubercle towards the posterior piriform cortex and amygdala (Supplementary Fig. 2a). Therefore, we counted axons corresponding to presumed mitral cells in the posterior aspect of the LOT using a confocal microscope (see Supplementary Fig. 2). We did not quantify the number of presumed tufted cells with this method owing to technical constraints: these individual axons were difficult to optically resolve due to the large number of cells labelled, the high intensity of labelling, their smaller diameter, and the more three-dimensional structure of the presumed tufted cell axon band in the LOT.

Image acquisition. Images were acquired on a Zeiss 710 confocal microscope (Zeiss) using a ×10 water immersion objective (Zeiss 0.45NA) or a Prairie in vivo microscope using a ×10 water immersion objective (Olympus 0.6NA). In both cases images were acquired in multiphoton mode using a Coherent laser tuned to 810 nm. Zen software with a custom-written macro was used for acquisition and tiling of images online; LSM 510 software was used for tiling in the event that it needed to be performed post-hoc (Zeiss). Images were also tiled using XUVTools[31] (Albert-Ludwigs-Universitat Freiburg). Images in which clearly labelled mitral cell fibres were not apparent in the LOT adjacent to the posterior piriform cortex were excluded from further study; in all such cases samples showed other signs of poor labelling (that is, poorly filled axonal arborizations).

Image alignment. During all image processing the operator was blind to the glomerular identity of the particular image stack. z-stacks of images of en bloc preparations of the cortical amygdala were aligned using an intensity-based, two-step linear–nonlinear protocol. Image alignment was carried out in Statistical Parametric Mapping 8 (SPM8) (available at http://www.fil.ion.ucl.ac.uk/spm), an open-source software package widely used in fMRI research. SPM is validated for a variety of registration tasks, including analysis of the axonal projection patterns of olfactory projection neurons in *Drosophila*[32]. The linear registration algorithm used a standard 12-parameter algorithm with Gauss-Newton optimization. The nonlinear registration modelled the spatial transformation with a set of basis functions, the discrete sine transformation (DST) and the discrete cosine transformation (DCT), of various spatial frequencies. Typically only a small number of nonlinear iterations were carried out (often less than 100) and the nonlinear regularization value was held relatively high (typically at 100, never below 1), which maximized the relative contribution of the linear alignment to the template. We generated a standardized template brain stack of the cortical amygdala by morphing four individual stacks onto a single high-quality image stack and then averaging the intensity of the resultant five individual stacks. This standardized reference stack was then used to warp all cortical amygdala stacks used in this study. Warping parameters were optimized for each image stack and the quality of warping was assessed by the overlap between the warped brain and the template image in the counterstained image channel only. The major landmarks in this preparation are the posterolateral cortical amygdala, the posteromedial cortical amygdala and the ventral edge of the posterior piriform cortex, all of which were used to judge alignment quality. Images that failed to align based on the counterstain were excluded from the data set. After warp parameters were defined based upon the counterstained channel, the channel containing the TMR-dextran staining was warped using identical parameters. Because after warping most of the data was contained in superficial planes, all analysis was performed on a maximum intensity projection of the warped volume. Alignment of maximal intensity projections of image stacks of the piriform cortex (which were highly anisotropic—that is, 20,000 × 10,000 × 40—precluding accurate volumetric alignment by SPM8) was performed in Photoshop using affine commands (such as zoom and rotate) as well as the "Warp" command, using only the counterstain as a guide to position.

Statistics. All error bars are ± standard error of the mean (s.e.m.). One-way MANOVA was used to test whether multivariate means of variables for different glomerulus types are different at the 0.05 significance level.

Image processing and cluster analysis. During all image processing the operator was blind to the glomerular identity of the particular image stack. Images were

processed by using a semi-supervised algorithm to filament trace the raw warped image stacks using the Imaris software package (Bitplane), slightly dilating the filaments, and then using those slightly dilated filaments to mask the original raw data. Before extraction of parameters, piriform and posterolateral nucleus projection patterns were aligned using the image alignment procedure previously described. Samples that were unable to be aligned due to poor counterstain or extreme physical distortion were excluded from k-means analysis. The data were then maximally contrasted so that differences in label intensity were normalized. A number of parameters characterizing the features of the axonal projection patterns to the piriform cortex (X and Y centre of mass coordinates, X and Y centroid coordinates, X coordinate of most anterior fibre, Y coordinate of most posterior fibre at LOT, number of TMR-positive pixels) and amygdala (X and Y centre of mass coordinates, X and Y centroid coordinates, X coordinate of medial-most fibre, absolute fibre density within the lateral and medial hemispheres of PLCo, ratio of lateral/medial fibre density to total fibre density within the PLCo) were extracted from the aligned and processed images of samples using ImageJ (NIH) and Matlab (Mathworks). No combination of parameters led to correct classification by glomerular type for piriform images, but the following parameters could be used in various combinations with nearly equal measures of success to classify amygdala samples: X coordinate of centre of mass, X coordinate of medial-most fibre, absolute fibre density within the medial hemisphere of PLCo, and the ratios of fibre density within the medial or lateral hemispheres to the total fibre density within the PLCo. The standardized Euclidean distance between pairs of objects in an M-by-N data matrix X, where rows of X correspond to the values of variables extracted from a single amygdala image and columns correspond to different variables, was then calculated using the "pdist" function in Matlab. The "linkage" function was then used to create a hierarchical cluster tree using the furthest distance between the clusters from the standardized Euclidean distance matrix calculated using "pdist". k-means clustering was performed using the "kmeans" function in Matlab, which partitions the points in the data matrix X into N clusters (for example, $N = 3$ for 3 glomeruli) by minimizing the sum, over all clusters, of the within-cluster sums of point-to-cluster centroid distances (squared Euclidean distance). The accuracy of this clustering is assessed by counting how many samples are correctly grouped together by glomerulus type. Calculation of the density of fibres was performed by generating a Z projection of the warped and processed images, and using the contour surface function of Imaris (Bitplane) to calculate the pixel density within the lateral and medial halves of the posterolateral cortical amygdala. The boundary between the lateral and medial halves of the poesterolateral nucleus was defined by the line extending straight horizontally through the semi-major axis of the posteromedial nucleus (which appears as an ellipse in the warped template brain; see Fig. 4d).

Normalized cross-correlation. Images of piriform cortex and posterolateral cortical amygdala projection patterns were blurred using a Gaussian filter (piriform cortex $\sigma = 200\ \mu m$; amygdala $\sigma = \sim 70\ \mu m$), cropped to exclude areas outside the region of interest, resized (piriform = 500×250 pixels; amygdala = 200×224 pixels), thresholded in Photoshop (Adobe) to binarize pixel values, and converted to grayscale in Matlab. The "normxcorr2" function in Matlab was then used to compute the normalized cross-correlation between the pixel values of two images (for example, MOR1-3 piriform 1 × MOR1-3 piriform 2) (Supplementary Fig. 8a). This function returns a matrix of Pearson correlation coefficients calculated using the following equation:

$$\gamma(u,v) = \frac{\sum x,y[f(x,y) - \bar{f}_{u,v}][t(x-u,y-v) - \bar{t}]}{\{\sum x,y[f(x,y) - \bar{f}_{u,v}]^2 \sum x,y[t(x-u,y-v) - \bar{t}]^2\}^{0.5}}$$

where f is the image, $\bar{t}$ is the mean of the template, and $\bar{f}_{u,v}$ is the mean of $f(x,y)$ in the region under the template.

By default, non-overlapping pixels in this analysis are assigned a padding value of zero; note that because the means of the template and the source image are subtracted from the pixel values, non-overlapping pixels can still potentially contribute to the correlation coefficient. This matrix of correlation coefficients is then plotted using the "surf" function in Matlab to generate correlograms (Supplementary Fig. 8c), where warmer colours represent high correlation values (close to 1) and cooler colours represent anticorrelation values (close to -1). The cross-correlation between pixel values in two images was calculated when the images were directly superimposed, and calculated again as one image was shifted relative to the other until no more overlap is possible in all directions (Supplementary Fig. 8b). The result of these calculations is a matrix of correlation coefficients that reflects the degree of correlation between the pixel values in two images across the entire range of shifts (until the images are completely non-overlapping), and this matrix can be plotted as a correlogram (Supplementary Fig. 8c). If there is similarity in the spatial patterning in two images, there will be a peak (red colour) in the correlogram that reflects correlation coefficients greater than zero. The

location of this peak in the correlogram is a direct reflection of the similarity between the spatial patterning in two images. Images with similar spatial patterning will have the highest correlation in pixel values when they are directly superimposed, and the result of this is a peak at the centre of the correlogram—the images do not need to be significantly shifted relative to each other for the spatial patterning to overlap. Images with dissimilar spatial patterning will have the highest degree of correlation in pixel values when one is shifted relative to the other (that is, shifting one image relative to the other gets the patterning in the images to overlap), resulting in a peak in the correlogram that is displaced from the centre. The amount the correlogram peak is spatially offset from the centre of the correlogram can be quantified, and used as a measure of the similarity of the spatial patterning in two images (Supplementary Figs 10, 11). Finally, the shape of the peak in the correlogram is related to the nature of the patterning; focal patterning will generate a more focal (circular) peak, whereas more distributed patterning will generate a more elongated peak (that is, there is overlap in the two patterns over a larger range of spatial shifts). For both piriform and amygdala, normalized cross-correlation was performed for all possible pairwise combinations, in all possible configurations (for example, MOR1-3 A × MOR28 A, and MOR28 A × MOR1-3 A). The location of the maximum correlation coefficient in correlograms was determined by transforming the matrix of correlation values into a linear array and using the "max" function to find the maximum value in the array, followed by the data cursor feature to find the X and Y coordinates of the maximum correlation coefficient in each correlogram. For each pairwise comparison, the location of the maximum correlation coefficient for the configuration where the distance between the location of the maximum correlation coefficient and the centre of the correlogram was the smallest was used to generate scatterplots.

Because non-overlapping pixels can potentially be assigned a real number value (due to subtracting of the means) in the calculation of the Pearson correlation coefficient, artefacts can be introduced in the correlogram at the edges, which represent correlation coefficients calculated under conditions where few pixels are overlapping between the template and the source images. We therefore also performed normalized cross-correlation analysis using an alternative data padding method to control for the possibility that the edge artefacts cause changes in the spatial position of the correlogram coefficient maximum. Template images tiled as a 3×3 array (Supplementary Fig. 8d) were generated for all individual images, and the normalized cross-correlation was then calculated for all pairwise combinations of single and tiled images, as previously described. Running the normXcorr2 algorithm using the tiled templates returns a correlogram that resembles a 3×3 grid with additional data points at the edges, but in which the block in the centre has been subject to analysis without the source image ever sliding off into a region that has no overlap. Because the template image is tiled, the region of overlap between the sliding image and the template will always contain the same set of pixel values, though the spatial order of these pixels will be offset. The result is that, for computing the centre block of the 3×3 cross-correlogram, the mean image pixel value and the standard deviation of image pixel values used to calculate the Pearson coefficient are constant as this region of the tiled image slides across the template; this effectively avoids the introduction of the edge artefacts described earlier. The resulting matrices of correlation coefficients were plotted using the "surf" function. These matrices were either left uncropped (and therefore include the correlations between all nine panels plus the edges), were cropped to the equivalent spatial displacements as generated in the initial analysis with zero padding (and which therefore contain offset peaks that are directly comparable between the two methods), or were cropped such that the source image only was allowed to slide for 50% of its length on any axis across the template image. This final crop limits all of the values in the correlation matrix to those in which half or more of the values in the correlation matrix arise from correlations between the image and the centre tile (rather than the adjacent tiles in the template). The X, Y coordinates of the maximum correlation coefficient were extracted from the "equivalent" correlograms using the "max" and "find" functions on a submatrix representing the correlation coefficient values within and surrounding the centre peak of the correlogram (for example, the peak closest to the centre of the correlogram, not the peaks near the edge of the correlogram in the equivalent view). Once again, for each pairwise comparison, the location of the maximum correlation coefficient for the configuration where the distance between the location of the maximum correlation coefficient and the centre of the correlogram was the smallest was used to generate scatterplots. The scatterplots generated using the two different methods are qualitatively and quantitatively similar (Supplementary Figs 10, 11 and Supplementary Tables 2, 3), and the correlograms generated by the two methods are qualitatively similar as well (Supplementary Fig. 12), suggesting that our analyses are robust to possible correlation artefacts caused by differences in the degree of pixel overlap between the two methods.

Genetics. Construction of MOR1-3–IRES–GFP animals was achieved as follows: a 7.8-kb genomic clone containing the complete MOR1-3 open reading frame

plus 5.3 kb upstream and 1.5 kb downstream sequence was mutagenized by PCR to insert a PacI restriction site 5 bp after the stop codon. An IRES–GFP–ACN cassette was cloned into the PacI site to prepare the targeting vector[33]. After electroporation into 129SvEv-derived mouse embryonic stem cells, genomic DNA from G418-resistant colonies was screened by Southern blotting with AflII to detect homologous recombination using a 400 bp probe 5′ of the targeting vector (7.4 kb wild-type allele versus 12.7 kb targeted allele). Chimaeras obtained from recombinant clones by standard mouse procedures were mated to C57BL/6J females to obtain heterozygous MOR1-3–IRES–GFP mice that had deleted the neo[R] selection marker in the male germline. Construction of the MOR174-9–IRES–GFP animals was achieved as follows: two genomic fragments containing the MOR174-9 5′ flanking sequence plus open reading frame (3.4 kb, XmaI sites) and 3′ flanking sequence (2.7 kb, SalI sites) were isolated by PCR from mouse genomic DNA, with restriction sites present in the primers. These two homologous arms were cloned into an IRES–GFP–ACN vector to obtain the targeting construct. Homologous recombinant clones were identified by EcoRI-digested genomic DNA blots (5.2 kb wild-type allele versus ~10 kb targeted allele) using a 300 bp probe 5′ of the construct. Mice were obtained from recombinant embryonic stem cells as described for MOR1-3–IRES–GFP strain. MOR28–IRES–GFP was previously described[34]. Note that MOR28 is also known as both MOR244-1 and Olfr1507; MOR1-3 is also known as Olfr66; and MOR174-9 is also known as Olfr73. The OMP–IRES–spH and M72–IRES–GFP animals were obtained from the Jackson Labs.

26. Haberly, L. B. & Price, J. L. The axonal projection patterns of the mitral and tufted cells of the olfactory bulb in the rat. *Brain Res.* **129,** 152–157 (1977).
27. Scott, J. W., McBride, R. L. & Schneider, S. P. The organization of projections from the olfactory bulb to the piriform cortex and olfactory tubercle in the rat. *J. Comp. Neurol.* **194,** 519–534 (1980).
28. Scott, J. W. Electrophysiological identification of mitral and tufted cells and distributions of their axons in olfactory system of the rat. *J. Neurophysiol.* **46,** 918–931 (1981).
29. Schneider, S. P. & Scott, J. W. Orthodromic response properties of rat olfactory bulb mitral and tufted cells correlate with their projection patterns. *J. Neurophysiol.* **50,** 358–378 (1983).
30. Skeen, L. C. & Hall, W. C. Efferent projections of the main and the accessory olfactory bulb in the tree shrew (*Tupaia glis*). *J. Comp. Neurol.* **172,** 1–35 (1977).
31. Emmenlauer, M. *et al.* XuvTools: free, fast and reliable stitching of large 3D datasets. *J. Microsc.* **233,** 42–60 (2009).
32. Datta, S. R. *et al.* The *Drosophila* pheromone cVA activates a sexually dimorphic neural circuit. *Nature* **452,** 473–477 (2008).
33. Bunting, M., Bernstein, K. E., Greer, J. M., Capecchi, M. R. & Thomas, K. R. Targeting genes for self-excision in the germ line. *Genes Dev.* **13,** 1524–1528 (1999).
34. Shykind, B. M. *et al.* Gene switching and the stability of odorant receptor gene choice. *Cell* **117,** 801–815 (2004).

LETTER

Sensory maps in the olfactory cortex defined by long–range viral tracing of single neurons

Sulagna Ghosh[1], Stephen D. Larson[2], Hooman Hefzi[1], Zachary Marnoy[1], Tyler Cutforth[3], Kartheek Dokka[1] & Kristin K. Baldwin[1]

Sensory information may be represented in the brain by stereotyped mapping of axonal inputs or by patterning that varies between individuals. In olfaction, a stereotyped map is evident in the first sensory processing centre, the olfactory bulb (OB), where different odours elicit activity in unique combinatorial patterns of spatially invariant glomeruli[1,2]. Activation of each glomerulus is relayed to higher cortical processing centres by a set of ~20–50 'homotypic' mitral and tufted (MT) neurons[3]. In the cortex, target neurons integrate information from multiple glomeruli to detect distinct features of chemically diverse odours[4–6]. How this is accomplished remains unclear, perhaps because the cortical mapping of glomerular information by individual MT neurons has not been described. Here we use new viral tracing and three-dimensional brain reconstruction methods to compare the cortical projections of defined sets of MT neurons. We show that the gross-scale organization of the OB is preserved in the patterns of axonal projections to one processing centre yet reordered in another, suggesting that distinct coding strategies may operate in different targets. However, at the level of individual neurons neither glomerular order nor stereotypy is preserved in either region. Rather, homotypic MT neurons from the same glomerulus innervate broad regions that differ between individuals. Strikingly, even in the same animal, MT neurons exhibit extensive diversity in wiring; axons of homotypic MT pairs diverge from each other, emit primary branches at distinct locations and 70–90% of branches of homotypic and heterotypic pairs are non-overlapping. This pronounced reorganization of sensory maps in the cortex offers an anatomic substrate for expanded combinatorial integration of information from spatially distinct glomeruli and predicts an unanticipated role for diversification of otherwise similar output neurons.

Chemical components of odours are detected by a large repertoire of olfactory sensory neurons (OSNs) based on their expression of one of ~1,400 odorant receptor (OR) genes (in the mouse)[7]. The axons of OSNs expressing the same OR coalesce in the olfactory bulb (OB) to form two glomeruli at mirror symmetric locations, where they form synapses with mitral and tufted (MT) neurons[2,8]. This produces a stereotyped topographic map in which different odours activate characteristic subsets of glomeruli and their associated MT neurons[1,9]. Odour recognition is thought to involve integration of inputs from different MT neurons, but the organization of glomerular inputs to cortical targets has not been described and whether sets of homotypic MT neurons are similar or diverse in cortical targeting is not known[4–6].

In other senses, such as vision, hearing and touch, adjacent neurons often respond to similar sensory stimuli and send axons to closely apposed cortical targets, producing sensory maps that are similar between individuals[10–12]. In contrast, anterograde and retrograde tracing of inputs to the primary olfactory cortex have shown that neighbouring MT neurons project to broad cortical regions and that small cortical regions receive input from distributed sets of MT neurons[13,14]. However, these studies did not examine neurons associated with a particular glomerulus. Therefore it is not clear whether the organization

of MT axons in the cortex retains aspects of the glomerular organization of the OB, or whether sensory maps in the cortex are stereotyped or divergent between individuals (Supplementary Fig. 1).

One barrier to mapping olfactory neural circuits is the lack of long-range axon tracing methods that do not obscure the glomerular identity of the MT neuron. We have developed a viral tracing system that overcomes these obstacles. This system exploits an attenuated strain of Sindbis virus to express high levels of three different 'colours' of fluorescent proteins in neurons after transduction with a single virion (Fig. 1a). This produces rapid labelling of MT dendrites and distant axonal branches without spillover effects at the injection site, allowing us to link neurons to a specific glomerulus and trace their projections with single-cell resolution, in three colours.

To determine whether MT neuron projections retain the stereotypic organization of the OB in different cortical targets, we labelled separate clusters of MT neurons with different coloured viral tracers and examined their axonal distributions in the anterior olfactory nucleus pars externa (AON pE) and the pyriform cortex (PC)
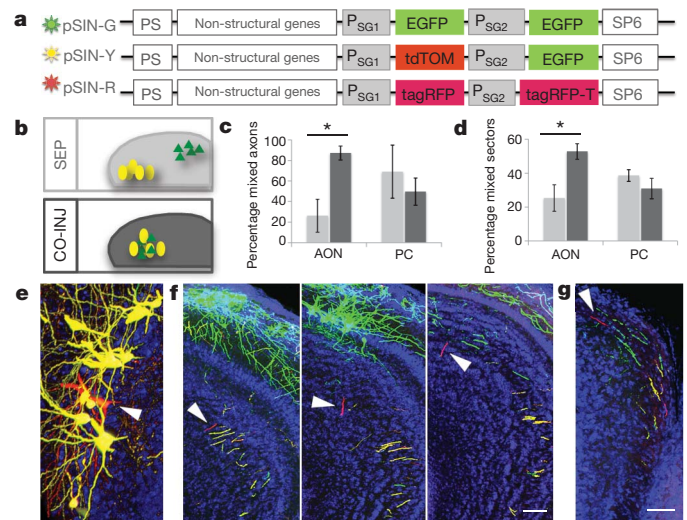

Figure 1 | Different organizations of MT axons in the AON and PC.
a, Sindbis virus vectors were altered so that infection with one virus produces two fluorescent proteins and results in green (pSIN-G), yellow (red+green, pSIN-Y) or red (pSIN-R) labelling. **b**, Schema of OB labelling with separate (SEP: light grey) or colocalized (CO-INJ: dark grey) injections. **c, d**, Segregation is maintained only in the AON based on analyses of confocal images from separate injections ($n = 4$ animals) and control co-injections ($n = 2$ animals) using two mixing indices (Student's t-test, $*P = 0.003$ (**c**), $*P = 0.03$ (**d**)). Error bars: 95% confidence. **e–g**, Some axons diverge from their neighbours as shown by tracing a red neuron (white arrows) that intermingles with yellow neurons in the OB (**e**) yet projects to a different region of AON pE which is primarily innervated by green axons from a distant OB region (**f, g**). Panels in **f** (left to right) are coronal OB sections, 400-μm apart. Blue, TOTO-3 nuclear staining. Scale bars, 100 μm. Panel **e** was rotated so lacks scale bars.

[1]Department of Cell Biology, Dorris Neuroscience Center, The Scripps Research Institute, La Jolla, California 92037, USA. [2]Program in Neurosciences, National Center for Microscopy and Imaging Research, University of California, San Diego, La Jolla, California 92037. USA. [3]Department of Molecular, Cell and Developmental Biology, University of California, Santa Cruz, California 95064, USA.

(Fig. 1b, Supplementary Fig. 2 and Supplementary Methods). In the AON pE, axons remained segregated, as previously reported[15,16]. Yet in the PC, the same axons become extensively intermingled, based on visual inspection and two quantitative indices of axon mixing (Fig. 1c, d, Supplementary Fig. 2 and Supplementary Methods). We speculate that the gross-scale spatial map in the AON pE may facilitate contralateral comparisons, whereas more complex patterns in the PC may permit increased combinatorial coding[6,17]. Intriguingly, in the AON pE, some MT axons diverge from their neighbours and intermingle with axons of distant MT cohorts, indicating that despite a gross-scale similarity, mapping in the AON pE differs from mapping in the OB at the single cell level (Fig. 1e–g).

Next, to determine how glomerular information is relayed to the AON and PC, we labelled individual MT neurons associated with a defined glomerulus. To accomplish this, we used mouse strains in which a stereotyped glomerulus is filled with green fluorescent protein (GFP) based on co-expression with a particular OR gene (*mOR174-9*, also known as *OR-EG* or *Olfr73*) (Fig. 2a, b). Tracing an individual MT neuron reveals local tufts in AON regions as well as extensive branching throughout the AONpE, AON pars principalis (pP) and anterior PC, with weaker signal in distal targets (Fig. 2c, Supplementary Fig. 4). This indicates that glomerular mapping by individual MT neurons is unlikely to be restricted to small selective cortical regions, which contrasts with the conclusions of studies in the rabbit that employed less sensitive tracers (Supplementary Methods and Supplementary Movie 1)[18].

A second new finding is that the axons of some MT neurons branch within the OB and innervate a small region of the granule cell layer, just proximal to the mirror symmetric glomerulus (Fig. 2b, d). This type of symmetric intrabulbar projection has been noted for external tufted (eT) neurons located in the glomerular layer, but not for neurons in the mitral cell layer[19]. Based on their morphology and internal projection,

we refer to these as internal tufted neurons (iT). In contrast with the eT neurons that project extensively to the olfactory tubercle, iT neurons project primarily to the AON and PC and do not exhibit detectable subtype-specific patterning (Supplementary Figs 3, 5 and 6)[20]. Consequently, iT and other MT neurons have been grouped in subsequent analyses.

To determine whether the complex patterns of MT axon branching and termination are stereotyped between individuals with respect to their glomerular input, we built a reference brain in which individual neuronal reconstructions could be aligned and mapped to different target regions (Supplementary Methods and Supplementary Fig. 4). Comparisons of homotypic EG-MT neurons ($n = 14$) and non-EG neurons ($n = 8$) revealed no detectable stereotypic patterns of branching or gross scale morphology (Fig. 3a–c and Supplementary Fig. 5).

Although the gross morphologies and branching patterns for homotypic OR-EG MT neurons are diverse, it remains possible that they exhibit stereotypy at axon termini, where synapses are likely to form. To address this possibility we aligned pairs of MT neurons in the reference brain and computed the distances between every axon terminal and its nearest neighbour in the comparison neuron. Similar analyses were performed using branch node locations to assess the closeness of fine distal branches. Control data sets show that this method detects similarities in branching and axon termination
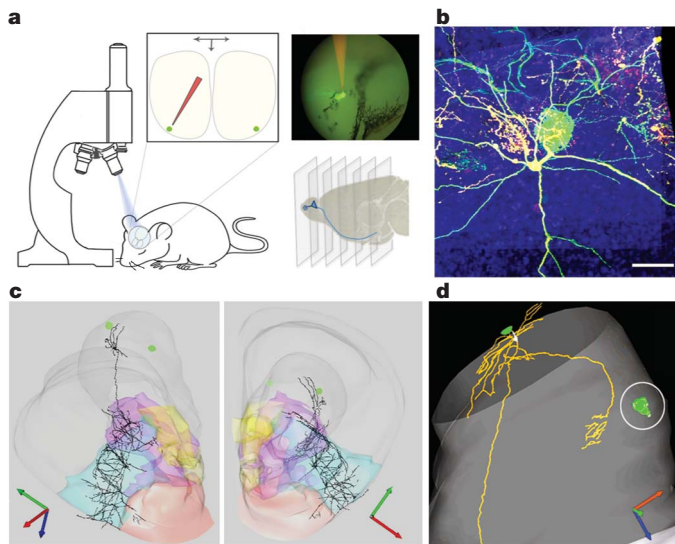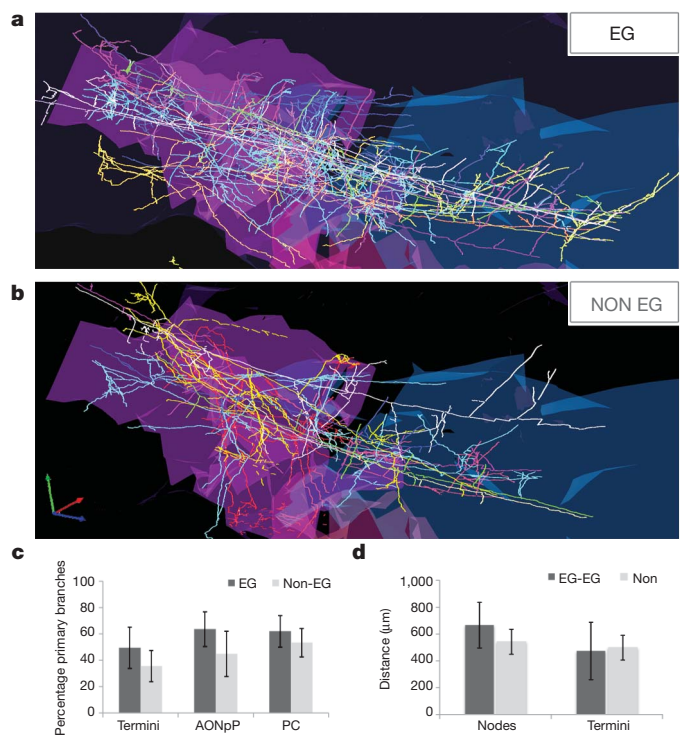

Figure 2 | Three-dimensional reconstructions of individual M/T neurons reveal extensive branching and intrabulbar mapping. a, Injections were targeted to the EG glomerulus *in vivo* using GFP (green in schema and fluorescence image, top right). Serial coronal sections were imaged using confocal scanning (lower right). **b,** Assignment of glomerular identity by tracing the primary dendrite of a neuron labelled with SIN-Y (yellow) into the EG glomerulus (green). Scale bar, 100 μm. **c,** Tracing axonal arbors with Neurolucida software produces a three-dimensional reconstruction in which branches are assigned to target regions. Green, EG glomeruli; purple, AON; cyan, PC; red, OT; yellow, taenia tecta. Blue axis, anterior–posterior; green axis, ventral–dorsal; red axis, medial–lateral. This neuron is from a 7-week-old mouse, all others are from 3-week-old mice. **d,** Three-dimensional reconstruction of the neuron in (**b**) reveals an intrabulbar axonal projection to the granule cell layer underlying the mirror symmetric EG glomerulus (white circle).

Figure 3 | Diverse patterns of glomerular inputs to the cortex. a, b, Fully traced EG ($n = 14$) and non-EG ($n = 8$) neurons were aligned to the reference brain to identify similarities between homotypic EG neurons (**a**), compared with non-EG neurons (**b**) (displayed traces were trimmed for clarity). Neurons branched extensively in the AON pE and pP (purple) and the anterior PC (cyan/blue). Blue axis, anterior–posterior; green axis, ventral–dorsal; red axis, medial–lateral (for **a** and **b**). **c,** Total primary axon branches in the dorsal or ventral direction were assigned to the AON pE, pP or aPC based on the reference brain. Primary branches that innervated a region (as percentage of total branches in that neuron) were averaged. No significant differences were observed between EG and non-EG neurons. Error bars, 95% confidence intervals. **d,** Proximity of axon termini and axon nodes were assessed for each pair of seven single MT neurons. Comparisons of homotypic EG-EG neuron pairs ($n = 12$) with heterotypic pairs (Non) ($n = 30$) revealed no differences in the average distance between nodes (**d**, left) or termini (**d**, right). Error bars, 95% confidence intervals, Student's *t*-test $P > 0.05$.

(Supplementary Fig. 7). However, homotypic EG-MT neuron pairs ($n = 6$) were no more similar than heterotypic pairs by any metric ($n = 15$) (Fig. 3d, Supplementary Methods and Supplementary Fig. 7).

Therefore, in contrast to other sensory systems, MT neurons ostensibly bearing similar sensory information do not branch or terminate in stereotyped positions. However, because these neurons were traced in several animals, differences in neural activity among animals or intrinsic differences in brain morphology may contribute to the observed variability. Alternatively, this diversity might be characteristic of homotypic neurons even in the same animal. To distinguish between these models, we injected multi-colour viral mixtures in the OR-EG region and identified sets of two to four neurons that could be traced to defined glomeruli, resulting in same-animal homotypic pairs ($n = 6$) and heterotypic pairs ($n = 7$) (Fig. 4a–d and Supplementary Fig. 8).

Same-animal analyses revealed extensive diversity among homotypic neurons. The majority ($\sim$70–90%) of primary branches among homotypic pairs are distinct in each target region (Fig. 4d, e). Homotypic pairs exhibit no more overlapping branches than do heterotypic pairs (Fig. 4d–f and Supplementary Fig. 8). Even when homotypic pairs seem to branch at the same point, fine-scale analyses reveal additional diversity. For example, in the AON pE two homotypic axons extend fine tufted collaterals in roughly the same region (Fig. 4b), yet heterotypic non-EG neurons project axonal tufts that interdigitate between the EG-MT axons (Fig. 4c).

Finally, it is possible that a subset of EG-MT neuron branches might display stereotypy, while others do not. To identify any overall patterns of similarity in EG-MT neurons that could reveal a partial stereotyped glomerular map in the cortex, we overlaid each possible pairing of matched homotypic EG-EG neurons ($n = 34$) and unmatched heterotypic sets of neurons ($n = 55$) and calculated the numbers of overlapping branches per pair, at three stringency levels. This revealed an increase in the average number of overlapping branches in the AON pP for homotypic OR-EG pairs. Whereas this result suggests that some stereotypy exists at the glomerular level in the AON pP, we note that this difference reflects only a modest overlap ($\sim$1.5 overlapping branches vs. $\sim$1) that is not significant at high stringency or when normalized to the total branches per region (Fig. 4f and Supplementary Methods).

Axon branching is an important determinant of the functional capacities of neural circuits, but the operational feature of a circuit is synaptic connectivity. Our studies cannot directly address this due to the lack of appropriate anterograde trans-synaptic tracing techniques. However, our characterization of axon ends reveals swellings at $\sim$90% of axon termini and varicosities in distal axon segments that have been suggested to represent sites of synapses (Supplementary Fig. 9)[15]. If axon termini indeed represent sites of synapses, our studies indicate that homotypic neurons form synapses with distinct sets of target neurons. For example, distances between nearest-neighbour axon ends in homotypic pairs often exceed 600 μm, which is larger than the reported dendritic fields for target neurons in the AON and PC ($\sim$200–500 μm)[21,22]. Recent retrograde transynaptic tracing studies provide direct support for this finding[23].

Here, we define the anatomy of the neural circuit that links odour detection in the nose with combinatorial processing in the brain. We show that diversity in the projections of MT neurons produces new cortical organizations of sensory information that differ from the OB in two ways. First, we observe extremely limited stereotypy in projections from a glomerulus, even in regions that maintain gross-scale stereotypic patterning. Second, the patterns of innervation of MT neurons are broadly distributed, suggesting that the set of $\sim$20–50 MT neurons are anatomically positioned to form synapses with a very large set of target neurons, which contrasts with the very precise patterns of inputs observed in the OB. This divergent, non-stereotyped organization contrasts with the highly stereotyped branching patterns of pairs of analogous neurons in the fly olfactory system. However, our findings concur with results of recent multi-neuron studies of the
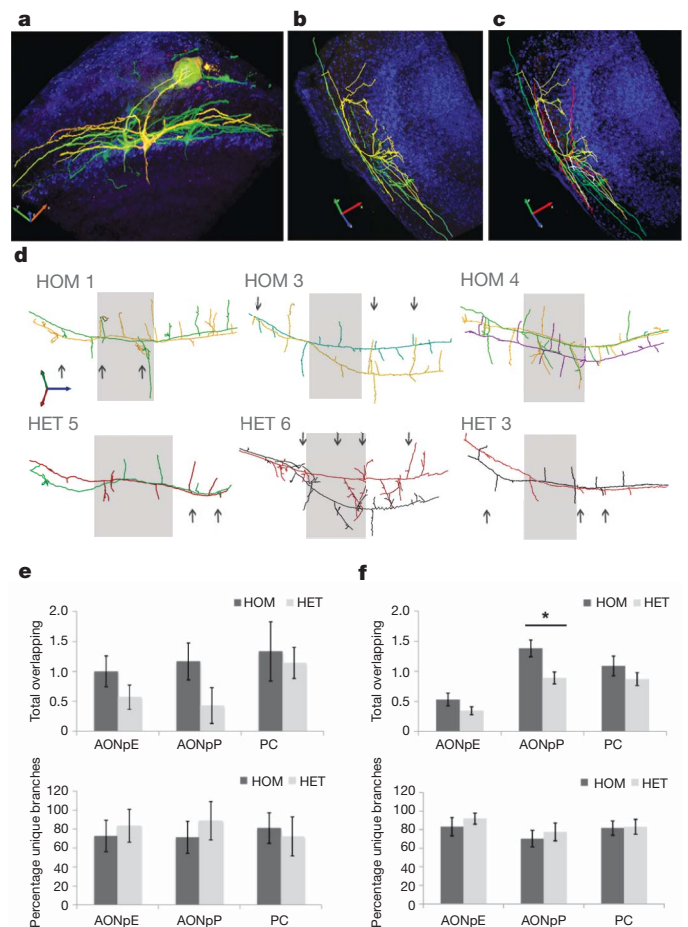

Figure 4 | **Extensive diversity among homotypic MT neurons in the same animal.** **a**, Viral labelling of a homotypic EG-MT neuron pair (collapsed confocal sections overlaid with traces). **b**, Homotypic axons in the AON pE exhibit distinct tufts. **c**, Two weakly-labelled non-EG M/T neurons (red, white) intercalate between the EG neurons (yellow, green). For **a–c**: yellow, SIN-Y; green, SIN-G; blue, nuclei. Blue axis, anterior–posterior; green axis, ventral–dorsal; red axis, lateral–median. **d**, Same-animal pairs of homotypic (HOM) and heterotypic neurons (HET) show diversity in primary branching. Overlapping branches (grey arrows) are shown. HOM 1 is the pair in **a**. Three EG neurons were labelled in one animal (HOM 4). Four neurons (two EG, two non-EG) were labelled in another animal (HOM 3/HET 3). Cortical regions: white left, AON pE; grey, AON pP; white right, PC. Main axon lengths vary between 1.9 and 2.7 mm, reflected by the width of the grey box. **e**, Even in the same animal, homotypic (HOM, $n = 6$) and heterotypic (HET, $n = 7$) neuron pairs are equally diverse in branch location (top) or percentage of unique primary branches per region (AONpE, AON pP and anterior PC; PC) (bottom). **f**, Compiling pairs from multiple animals ($n = 34$ HOM; $n = 55$, HET) revealed increased branch overlap in homotypic neurons in the AON pP (top), which is not significant when normalized to the total branches in that region (bottom). Error bars show s.e.m. (top) or s.d. (bottom) Significance was assessed by a Mann–Whitney U Test (*$P = 0.008$) and by Student's t-test adjusted for multiple comparisons (*$P = 0.004$, for $P < 0.05$, $P_{adj} = 0.016$).

mouse olfactory cortex, and with single neuron tracing in the zebra-fish[23–26]. We speculate that the increased numbers and diversity of homotypic output neurons in vertebrate systems serve to enhance the capacity of cortical targets to perform combinatorial integration.

With respect to developmental mechanisms, this cortical reorganization presents a puzzle. In other sensory systems, neighbouring projection neurons send axons to adjacent target regions on the basis of similarities in molecular cues that derive from their spatial location and/or their patterns of neural activity. In the olfactory system, homotypic MT cells reside in similar locations in the OB and display highly

correlated sensory evoked activity[27–29]. Yet, we observe that homotypic MT neurons are highly diverse in their cortical projections. This observation is hard to reconcile with determined patterning mechanisms based on activity or molecular cues. Yet the alternative, that the selection of synaptic targets is stochastic, would predict that individuals possess different cortical representations of the same odour. Our method for mapping axonal projections of defined sets of olfactory neurons will enable more mechanistic studies of olfactory neural circuit formation and function, and may be useful for studies of other neural circuits that involve long-range connectivity.

METHODS SUMMARY

Recombinant Sindbis viruses were generated with standard cloning techniques. Viruses were injected into GFP-positive EG glomeruli using a fluorescence microscope (SZX12, Olympus). After 48 h, brains were sectioned coronally and all sections were imaged (Olympus FV500 laser scanning confocal microscope). Neuronal reconstructions were generated by manually tracing labelled axons in Neurolucida (v9.13, MicroBrightField). The reference brain, generated from an age- and genotype-matched animal, was imaged and olfactory regions were annotated using the Allen Reference Atlas and the contour map function in Neurolucida. The Whole Brain Catalog (version 0.7.7.8, http://wholebraincatalog.org) software platform permitted manipulation of traced neurons in three-dimensions for alignments to landmarks including EG glomeruli and identifiable clusters of nuclei in the AON, PC and OT.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Rubin, B. D. & Katz, L. C. Optical imaging of odorant representations in the mammalian olfactory bulb. *Neuron* **23,** 499–511 (1999).
2. Vassar, R. *et al.* Topographic organization of sensory projections to the olfactory bulb. *Cell* **79,** 981–991 (1994).
3. Haberly, L. B. & Price, J. L. The axonal projection patterns of the mitral and tufted cells of the olfactory bulb in the rat. *Brain Res.* **129,** 152–157 (1977).
4. Apicella, A., Yuan, Q., Scanziani, M. & Isaacson, J. S. Pyramidal cells in piriform cortex receive convergent input from distinct olfactory bulb glomeruli. *J. Neurosci.* **30,** 14255–14260 (2010).
5. Lei, H., Mooney, R. & Katz, L. C. Synaptic integration of olfactory information in mouse anterior olfactory nucleus. *J. Neurosci.* **26,** 12023–12032 (2006).
6. Stettler, D. D. & Axel, R. Representations of odor in the piriform cortex. *Neuron* **63,** 854–864 (2009).
7. Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65,** 175–187 (1991).
8. Ressler, K. J., Sullivan, S. L. & Buck, L. B. Information coding in the olfactory system: evidence for a stereotyped and highly organized epitope map in the olfactory bulb. *Cell* **79,** 1245–1255 (1994).
9. Khan, A. G., Parthasarathy, K. & Bhalla, U. S. Odor representations in the mammalian olfactory bulb. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2,** 603–611 (2010).
10. Petersen, C. C. The functional organization of the barrel cortex. *Neuron* **56,** 339–355 (2007).
11. Schreiner, C. E. & Winer, J. A. Auditory cortex mapmaking: principles, projections, and plasticity. *Neuron* **56,** 356–365 (2007).
12. White, L. E. & Fitzpatrick, D. Vision and cortical map development. *Neuron* **56,** 327–338 (2007).
13. Buonviso, N., Revial, M. F. & Jourdan, F. The projections of mitral cells from small local regions of the olfactory bulb: an anterograde tracing study using PHA-L (*Phaseolus vulgaris* leucoagglutinin). *Eur. J. Neurosci.* **3,** 493–500 (1991).
14. Luskin, M. B. & Price, J. L. The distribution of axon collaterals from the olfactory bulb and the nucleus of the horizontal limb of the diagonal band to the olfactory cortex, demonstrated by double retrograde labeling techniques. *J. Comp. Neurol.* **209,** 249–263 (1982).
15. Scott, J. W., Ranier, E. C., Pemberton, J. L., Orona, E. & Mouradian, L. E. Pattern of rat olfactory bulb mitral and tufted cell connections to the anterior olfactory nucleus pars externa. *J. Comp. Neurol.* **242,** 415–424 (1985).
16. Yan, Z. *et al.* Precise circuitry links bilaterally symmetric olfactory maps. *Neuron* **58,** 613–624 (2008).
17. Kikuta, S. *et al.* Neurons in the anterior olfactory nucleus pars externa detect right or left localization of odor sources. *Proc. Natl Acad. Sci. USA* **107,** 12363–12368 (2010).
18. Ojima, H., Mori, K. & Kishi, K. The trajectory of mitral cell axons in the rabbit olfactory cortex revealed by intracellular HRP injection. *J. Comp. Neurol.* **230,** 77–87 (1984).
19. Belluscio, L., Lodovichi, C., Feinstein, P., Mombaerts, P. & Katz, L. C. Odorant receptors instruct functional circuitry in the mouse olfactory bulb. *Nature* **419,** 296–300 (2002).
20. Nagayama, S. *et al.* Differential axonal projection of mitral and tufted cells in the mouse main olfactory system. *Front. Neural Circuits* **4** 1–12 (2010).
21. Brunjes, P. C. & Kenerson, M. C. The anterior olfactory nucleus: quantitative study of dendritic morphology. *J. Comp. Neurol.* **518,** 1603–1616 (2010).
22. Poo, C. & Isaacson, J. S. Odor representations in olfactory cortex: ''sparse'' coding, global inhibition, and oscillations. *Neuron* **62,** 850–861 (2009).
23. Miyamichi, K. *et al.* Cortical representations of olfactory input by trans-synaptic tracing. *Nature* advance online publication, doi:10.1038/nature09714 (22 December 2010).
24. Wong, A. M., Wang, J. W. & Axel, R. Spatial representation of the glomerular map in the *Drosophila* protocerebrum. *Cell* **109,** 229–241 (2002).
25. Miyasaka, N. *et al.* From the olfactory bulb to higher brain centers: genetic visualization of secondary olfactory pathways in zebrafish. *J. Neurosci.* **29,** 4756–4767 (2009).
26. Sosulski, D. L. *et al.* Distinct representations of olfactory information in different cortical centres. *Nature* advance online publication, doi:10.1038/nature09868 (30 March 2011).
27. Chen, T. W., Lin, B. J. & Schild, D. Odor coding by modules of coherent mitral/tufted cells in the vertebrate olfactory bulb. *Proc. Natl Acad. Sci. USA* **106,** 2401–2406 (2009).
28. Christie, J. M. *et al.* Connexin36 mediates spike synchrony in olfactory bulb glomeruli. *Neuron* **46,** 761–772 (2005).
29. Dhawale, A. K., Hagiwara, A., Bhalla, U. S., Murthy, V. N. & Albeanu, D. F. Non-redundant odor coding by sister mitral cells revealed by light addressable glomeruli in the mouse. *Nature Neurosci.* **13,** 1404–1412 (2010).

Author Contributions S.D.L. developed code to transfer and align neuron traces in the WBC platform. H.H. performed sector and proximity analyses. Z.M., S.G. and K.D. generated the three-dimensional reconstructions of neurons. K.D. assisted generating the Reference Brain. T.C. generated the mOR174-9-GFP mouse strain and edited the manuscript. S.G. designed and performed experiments, analysed data and edited the manuscript. K.K.B. conceived of the experimental design, analysed data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.K.B. (kbaldwin@scripps.edu).

METHODS

Viral constructs. All Sindbis virus constructs were generated using a modified pSinRep5 (Invitrogen) vector containing two subgenomic promoters (pSin2gene-EGFP, a gift from G. Patrick). pSIN-G and pSIN-Y were generated by cloning the coding sequence of either EGFP (Clontech) or tDTomato[30] into the Xba- and Sph-cut sites of the pSIN2gene-EGFP vector backbone. To generate pSIN-R, tRFP[31] and tRFP-T[30] were inserted into the vector using Xba/Sph- and BssHII/Mlu-cut sites in a pSIN-G vector that was modified to contain these sites by addition of a polylinker.

Production of Sindbis virus. For the production of Sindbis virus tracers, recombinant RNA and helper DH6 RNA were transcribed using the SP6 mMessage mMachine Kit (Ambion), and electroporated into BHK cells using a BTX ECM 600 at 220 V, 129 Ω and 1,050 μF. Virus was collected after 16–18 h and concentrated by centrifugation (20–30 min, 3,220g, 4 °C) through 7 ml/9K Pierce Concentrators (Thermo Scientific). The concentrated virus was stored in aliquots at −80 °C until further use.

Mouse strains. ORT mice were generated by crossing two previously published mouse strains with an additional strain. In this strain, an expression cassette containing an internal ribosome entry site (IRES) preceding a reporter gene (either *EGFP* or *TauLacZ*) was inserted into an odorant receptor gene immediately following the stop codon. The three lines were bred to homozygosity for each gene. The mouse lines are P2-IRES-TauLacZ[32], OR-174-IRES-GFP (T.C., unpublished) and MOR28-IRES-EGFP[33].

***In vivo* injection of Sindbis virus tracers.** All experiments were conducted on 3-week-old mice with the exception of a 6-week-old old litter used for comparison and assembly of the initial neuronal reconstruction. Mice were anesthetized with isoflurane (3% in 100% O_2) and placed in a stereotaxic apparatus. The skull was exposed and GFP-labelled glomeruli were visualized through the bone using a fluorescence stereo microscope (Olympus SZX12). A small open-skull cranial window was created above the glomeruli and viral tracers were injected directly into the targeted glomerulus using a glass micropipette ($\sim$10 μm tip diameter) attached to a customized manipulator (Narishige) and a high-pressure microinjector (Picospritzer). The virus tracer mix consisted of concentrated virus (SIN-G = 4×10^3 infectious units per microlitre (μl), SIN-Y = 3×10^3 infectious units per μl) in BHK media tinted with 0.1% Fast Green dye solution to aid visualization of injection. For double-labelling experiments, a mixed preparation of SIN-G, SIN-Y and/or SIN-G was used. Pressure and injector settings were optimized to deliver nanovolumes of virus mix which consistently labelled between 1 and 10 neurons on average per sample (30 p.s.i., 4 ms, 4–8 pulses). Mice were euthanized 48 h after injection. All procedures were performed in accordance with the guidelines and standards of the Scripps Research Institutes' Animal Care and Use Committee.

Tissue preparation. Injected animals were euthanized with isoflurane and perfused intracardially with PBS and 4% PFA. Brains with intact olfactory bulbs were isolated and post-fixed overnight at 4 °C. The olfactory bulb was sectioned along the coronal plane at 80 μm using a vibrating microtome (Leica VT1000S) and scanned for the presence of labelled MT neurons through an upright fluorescence microscope (Olympus BX60). MT neurons were identified by the laminar location of their soma and the presence of a primary dendrite innervating a single glomerulus. Cortices of selected samples with brightly labelled neurons were sectioned at 50 μm. Sections were labelled with TOTO-3 (Invitrogen, Molecular Probes) nuclear stain to delineate cortical areas and mounted using Gelvatol. The sequential order of sections was maintained throughout the entire procedure.

For the reference atlas, a 3-week-old male was perfused as above and the brain was post-fixed overnight at 4 °C. The brain was sectioned at 100 μm, stained and mounted as described above. The right hemisphere was used for imaging and mapping.

Imaging and image processing. Samples with labelled neurons were imaged sequentially at ×20 (PlanAPO oil, 0.8 NA) with a step size of 1.2 μm or 1.8 μm using an Olympus FV500 confocal microscope equipped with 488/543/633 nm excitation lasers and 505–525/560-600/660 nm emission filters. To aid with alignment of neuron traces, ×4 (UPlanFl 0.13 NA) images of the entire brain and ×10 (UplanAPO oil, 0.4 NA) images of the bulb were also acquired. Metamorph and Fluoview software were used to obtain maximum intensity projections of confocal stacks.

Mosaic imaging of the reference brain was carried out at ×10 (UPlanSApo NA 0.4) with a step size of 50 μm using an Olympus FV1000. The reference brain images per tissue section were tiled and collapsed using Image J (NIH) plugin MEMontage (NCMIR).

In Adobe Photoshop, maximum intensity projections of sample images were digitally enhanced using levels, curves, brightness and contrast primarily for display purposes. The red, yellow and green axons in Fig. 1f, g, Supplementary Fig. 2a, b and Supplementary Fig. 4a were selectively enhanced to improve visualization in three-colour printed version of this manuscript in order to accurately represent the data that is apparent in the individual channel representations of each stack. The raw images for Supplementary Fig. 2a, b separated by channel and minimally enhanced using levels are provided for comparison in Supplementary Fig. 2c, d.

Construction of 3-week-old mouse Reference Brain Atlas. Tiled maximum intensity projections of the reference brain sections were imported into Neurolucida software (Version 9.13, MicroBrightfield) and aligned using distinct nuclear layers and section outlines. Only rigid body transformations were used to sequentially align the images along the antero-posterior axis with a fixed z-distance of 100 μm between each section. The Allen Reference Atlas was used to annotate the different olfactory cortical regions (MOB, LOT, AON, PC, dTT, vTT, OT) and outlines of each region were traced using the contour map function in Neurolucida as per the manufacturer's instructions. These contours were exported from Neurolucida in their XML format and converted into triangle meshes, one per brain region, using a MicroBrightfield converter developed at the National Center for Microscopy and Imaging Research (NCMIR), available online at http://sourceforge.net/projects/ccdb-support/. The Whole Brain Catalog (development version 0.7.7.8, http://wholebraincatalog.org) software platform was used to visualize and manipulate the reference brain in three-dimensional space.

Neuronal reconstruction and three-dimensional alignment. Unmodified confocal images of single or double neuron samples were imported into Neurolucida and aligned using the main axon and distant collaterals. Axon tracing was carried out as per the manufacturer's instructions. To visualize each neuronal trace in the context of the olfactory cortex, two different techniques were used. First, for a small subset of the neurons, ×4 images of the corresponding brain sections were used to contour map cortical regions alongside the trace to create a complete three-dimensional reconstruction of the neuron within the source brain. Second, a less precise but more high-throughput technique was employed to map all the traces in the context of the three-dimensional reference brain. Comparable fiducial markers such as labelled glomeruli and distinct nuclei in the AON, PC and OT (Allen reference brain slides 21, 25, 31, 34) were manually identified for each sample. Tiled ×10 maximum intensity projections of the entire brain section containing these markers were obtained for each neuron sample. In Neurolucida, trimmed (main axons with collaterals spanning at least 50 μm from the node) or complete representations of each neuron trace along with its set of image sections were exported in XML format. The MicroBrightfield converter described above with an adaptation provided by the Whole Brain Project team (http://code.google.com/p/wholebrain/source/browse/#svn) allowed for the conversion of XML files to NeuroML format (http://neuroml.org). This code also allowed the conversion of the image metadata in the Neurolucida XML into the Whole Brain Catalog format for representing images in three-dimensional space. Further, the development version 0.7.7.8 of the Whole Brain catalogue software platform was augmented to allow the use of handles to move groups of two-dimensional images and three-dimensional neuronal morphologies (the 'data sets') as a unit. This functionality gave us the ability to align multiple neuronal reconstructions in the context of the reference brain meshes and move them independently with simple rotation and translation operations. The two-dimensional images with fiducial markers were used to align the data sets with the reference brain boundaries. The AP axis was used as the primary axis for alignment and the traces were rotated to achieve optimal fit. This ensured that the three-dimensional traces would be approximately in place relative to one another, and the reference brain, to facilitate the analysis of overall targeting specificities and extent of overlap between individual neurons.

Data collection and statistical analysis. Sector and proximity analysis: two independent assays were used to analyse the multicolour multi-neuron data set. For the sector analysis, maximum intensity projections of four analogous regions spanning approximately 400 μm in the AON and PC from five different mice were aligned flush against the top left corner in Photoshop. Using the grid tool, 128-pixel square sectors were marked on the image. The sector size was chosen such that axonal segments on average traversed over more than one sector. Each sector with main axons within it was classified as unmixed (axons of one colour within) or mixed (axons of multiple colours within). For samples in which SIN-G and SIN-Y had been injected into separate locations (S1, S2, S4) or co-injected (CO1), this resulted in bins of unmixed green sectors and yellow sectors, or mixed yellow-green sectors. In S3/CO3 samples, a mixture of SIN-G and SIN-Y was injected in a distinct region from SIN-R. Mixing of green and yellow fibres was reported as CO3. Mixing or segregation of either green or yellow with red was reported at S3. This allowed us to compare the relative positions of segregated and co-localized inputs in the same animal across multiple regions. In the sector analysis for S3/CO3, sectors were classified as containing only green or yellow axons, only red axons, or mixed (having green or yellow and red axons). In addition, as an internal control SEG3/MIX3 was further analysed by counting unmixed sectors containing only green axons or only yellow axons and with mixed being defined as containing green and yellow axons (ignoring red axons). Ratios of mixed/total ('mixing

index') and unmixed/total sectors were plotted to normalize for the number of sectors covered by each type of axon. To determine whether the topography in the bulb is preserved in the AONpE and the PC, two-tailed Students *t*-test was used to compare the extent of mixing in separate versus co-injected samples.

In the proximity analysis method, the main axon of each green, yellow and red (for S3/CO3) neuron was reduced to a single point on the maximum intensity projection, with attached coordinates. The single point corresponded to the approximate midpoint of each axon segment, which was selected by manual measurement of the axon segments in collapsed image sections. The distance between each axon and every other axon in the section was computed using a customized Pdist function in Matlab. To determine whether the axons nearest to each other tended to be of similar or different origin (or colour) we calculated the maximum distance to any axon (providing an internal control that normalized variations in region size distribution of injections and the position of the axon in the region). Next, for each axon we determined how many neighbouring axons were of the same colour (as a percentage of total axons). This normalized for differences in the number of neighbours of different neurons.

Next, we identified the axons that were close (defined at 15% of the maximum distance). The relative percentage of same colour/ different colour was computed and if this exceeded 80%, this axon was called segregated. Other axons were called mixed. The cumulative number of segregated versus mixed axons across multiple sections from the AON and PC were compiled for five animals, with MIX3 and SEG3 being green versus yellow, and green/yellow versus red as before. To assess the robustness of this analysis, we performed 12 different computational analyses of these regions, varying the values for proximity (from 12–18%) and for segregation (from 70–80%). All analyses gave similar results. We present a representative yet stringent data set. Significance was assessed by two-tailed Student's *t*-test.

Single neuron analysis: neurons selected for this tracing study were distinguished from other cells in the olfactory bulb based on morphological features such as shape of cell soma, laminar location of cell body, primary dendritic target and the absence or presence of axon collaterals within the bulb. For each neuron labelled in the MT cell layer we partially reconstructed its dendritic and axonal projections in the olfactory bulb. Cells with a slightly displaced tubular soma and distinct intrabulbar axon collaterals that projected to the mirror symmetric glomerulus were classified as internal tufted neurons. Cells with distinct pyramidal soma and no bulbar collaterals were categorized as mitral type I neurons. Cells in the mitral cell layer with intermediate morphologies and varying degrees of intrabulbar axon collaterals were classified as mitral type II neurons. A small number of cells were binned as unclassified due to ambiguities in assigning cell type identity. These cells were not included in cell type specific analyses.

For the single neuron branch analysis and degree of overlap, individual neurons were aligned in the context of the reference brain as described previously. The number of primary branches extended by each neuron in the AON pE, AON pP, and aPC was recorded in a blind format. Bidirectional branches were counted as being both dorsal and ventral. Differences in branching within the three brain regions were assessed using Student's *t*-test (two groups) with a Bonferroni correction for multiple comparisons ($n = 6$ for branch number). The extent of overlap between neurons was analysed first in neuron pairs from the same animal and then extended to a pair-wise comparison of multiple neurons from different animals.

For each pair the numbers of overlapping primary branches along the dorsal and ventral axes in each of the three cortical regions were counted. Three degrees of stringency (relaxed, moderate and strict) were used to define the extent of overlap between neurons based on the necessity for additional warping or transformation to accommodate a better fit. Data reported did not differ significantly between the moderate and relaxed counts; however, the overlap in the AON pP lost significance under the strict analysis. Each set of comparisons was analysed by Student's *t*-test with a Bonferroni correction for multiple comparisons ($n = 3$ for percent overlap). Significance was reported for the adjusted number based on a pre-correction significance of $P < = 0.05$. We also performed a Mann–Whitney U test for ranked data exhibiting non-normal distribution, which confirmed the significance of the effects reported.

Terminal and node proximity analyses: to compare the distribution of nodes and termini of neurons with respect to one another, we performed a nearest neighbour analysis of axonal arbors. To accomplish this we extended the Whole Brain Catalog client libraries written in Java to be accessible via the Python programming language by means of the Jython project (http://jython.org). We made use of the WBC's integration with the Java Universal Network/Graph framework (http://jung.sourceforge.net) to convert the NeuroML structure of the neurons into a tree data structure that enabled graph operations such as identifying all endpoints or nodes efficiently. Once the correct set of points were identified, their x, y and z coordinates were read into a 'point matrix' that had three columns (one per coordinate) and n rows, where n was equal to the number of points to be analysed. We then used the JNumerics package via Jython to compare the 'point matrix' of different neurons to each other. In performing the nearest neighbour analysis, we made pairwise comparisons. For each row (representing a point) in point matrix A, we calculated the Euclidean distance to every other row in point matrix B. This resulted in an n by m 'distance matrix' ($n =$ no. of rows in A, $m =$ no. of rows in B). For each row in the 'distance matrix' we sorted the distances and chose the smallest value, which represents the distance to the nearest neighbour for each point. We then used the resulting 'nearest-neighbour distance vector' to plot a 10-bin histogram for each pair-wise comparison in our set of seven fully traced neurons. We used the Python NumPy and MatPlotLib libraries to carry out the plotting. The resulting scripts have been made available as part of the International Neuroinformatics Coordinating Facilities Object Model for Neuroinformatics (http://incf-omni.googlecode.com).

The validity of this assay was tested using control pairs of artificially aligned neurons with predictable outcomes. The distribution plots generated by Matlab were then characterized into two distinct classes based on the location of the peak bin. Neuron pairs with a skew value of greater than 1 (peak bin towards minima) were assigned to class I and those with a skew value of less than 1 (peak bin towards maxima) were assigned to class II.

30. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature Biotechnol.* **22,** 1567–1572 (2004).
31. Merzlyak, E. M. *et al.* Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nature Methods* **4,** 555–557 (2007).
32. Mombaerts, P. *et al.* Visualizing an olfactory sensory map. *Cell* **87,** 675–686 (1996).
33. Shykind, B. M. *et al.* Gene switching and the stability of odorant receptor gene choice. *Cell* **117,** 801–815 (2004).

LETTER

Recapitulation of premature ageing with iPSCs from Hutchinson–Gilford progeria syndrome

Guang-Hui Liu[1], Basam Z. Barkho[1], Sergio Ruiz[1], Dinh Diep[2], Jing Qu[1], Sheng-Lian Yang[1], Athanasia D. Panopoulos[1], Keiichiro Suzuki[1], Leo Kurian[1], Christopher Walsh[1], James Thompson[3], Stephanie Boue[4], Ho Lim Fung[2], Ignacio Sancho-Martinez[1], Kun Zhang[2], John Yates III[3] & Juan Carlos Izpisua Belmonte[1,4]

Hutchinson–Gilford progeria syndrome (HGPS) is a rare and fatal human premature ageing disease[1–5], characterized by premature arteriosclerosis and degeneration of vascular smooth muscle cells (SMCs)[6–8]. HGPS is caused by a single point mutation in the lamin A (*LMNA*) gene, resulting in the generation of progerin, a truncated splicing mutant of lamin A. Accumulation of progerin leads to various ageing-associated nuclear defects including disorganization of nuclear lamina and loss of heterochromatin[9–12]. Here we report the generation of induced pluripotent stem cells (iPSCs) from fibroblasts obtained from patients with HGPS. HGPS-iPSCs show absence of progerin, and more importantly, lack the nuclear envelope and epigenetic alterations normally associated with premature ageing. Upon differentiation of HGPS-iPSCs, progerin and its ageing-associated phenotypic consequences are restored. Specifically, directed differentiation of HGPS-iPSCs to SMCs leads to the appearance of premature senescence phenotypes associated with vascular ageing. Additionally, our studies identify DNA-dependent protein kinase catalytic subunit (DNAPKcs, also known as PRKDC) as a downstream target of progerin. The absence of nuclear DNAPK holoenzyme correlates with premature as well as physiological ageing. Because progerin also accumulates during physiological ageing[6,12,13], our results provide an *in vitro* iPSC-based model to study the pathogenesis of human premature and physiological vascular ageing.

Three HGPS primary fibroblast lines, originally isolated from patients with the classical *LMNA* mutation (Gly608Gly), were transduced with retroviruses encoding OCT4, SOX2, KLF4, c-MYC and green fluorescent protein (GFP). NANOG-positive colonies were effectively obtained when using early passage, but not late passage (>25), HGPS fibroblasts (Supplementary Fig. 1a). We focused on iPSC lines of a well characterized HGPS fibroblast line, AG01972 (refs 9–12). Compared to normal fibroblasts, HGPS fibroblasts (AG01972) showed abnormal nuclear morphology, reduced expression of the lamina components lamin B1 and LAP2β (also known as TMPO), loss of heterochromatin markers H3K9me3, HP1α (also known as CBX5) and HDAC1, and reduced expression of nuclear proliferation marker Ki67 (also known as MKI67; Fig. 1a and Supplementary Fig. 2). From HGPS fibroblasts, we derived six iPSC lines. In addition, we generated control iPSC lines from wild-type fibroblasts (BJ and IMR-90 cell lines). Control and HGPS iPSC lines demonstrated pluripotent gene expression, demethylation of the OCT4 promoter and transgene silencing (Fig. 1b and Supplementary Figs 1, 3 and 4a, and data not shown). They were maintained for more than 50 passages without a loss of pluripotency or the acquisition of detectable morphological or growth abnormalities. The pluripotency of each iPSC line was assessed by differentiation into the three embryonic germ layers *in vitro*, using embryoid body (EB) formation, and/or *in vivo*, by teratoma formation (Supplementary Fig. 5a–d). Out of these lines, we focused primarily on

HGPS-iPSC#4 and BJ-iPSC#3 for our studies (hereafter referred to as HGPS-iPSC or BJ-iPSC). Both BJ-iPSCs and HGPS-iPSCs were able to differentiate towards specialized mesoderm-derivatives such as smooth muscle cells (SMC), endothelial cells (Supplementary Fig. 6a), or beating cardiomyocytes (Supplementary Movies 1 and 2). Moreover, all analysed iPSC lines showed normal chromosomal integrity (Supplementary Fig. 5e). Finally, *LMNA* sequencing confirmed the presence of the classical mutation in HGPS-iPSCs (Supplementary Fig. 5f). Altogether, these data indicate that the somatic cells from HGPS patients, despite their significant premature senescence phenotypes and nuclear defects, have been properly reprogrammed and can be effectively maintained in a pluripotent state.

Lamin A/C protein is expressed in differentiated somatic cells but is absent in embryonic stem cells (ESCs)[11,14]. Therefore, we next examined the expression of lamin A/C in the generated iPSC lines. As shown in Fig. 2a, lamin A/C expression is significantly downregulated in iPSCs, compared to their parental fibroblasts, whereas lamin B1 transcripts were upregulated. Although progerin should follow a similar pattern of expression as observed for lamin A/C, *LMNA* expression is independent
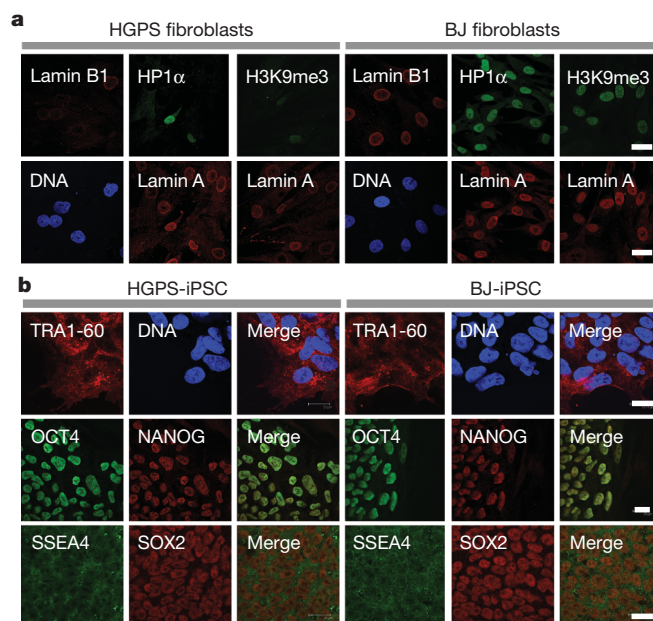

Figure 1 | Generation of iPSCs from HGPS fibroblasts.
a, Immunofluorescence analysis performed on HGPS (left) and BJ (right) fibroblasts at passage 17 with the indicated antibodies. **b**, Immunofluorescence analysis of the indicated pluripotent markers in HGPS-iPSCs (left) and BJ-iPSCs (right). Nuclei were visualized with Hoechst stain (blue). Scale bar, 20 μm.

[1]Gene Expression Laboratory, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. [2]Department of Bioengineering, University of California at San Diego, La Jolla, California 92093, USA. [3]Department of Cell Biology, Scripps Research Institute, La Jolla, California 92037, USA. [4]Center for Regenerative Medicine in Barcelona, Dr. Aiguader 88, 08003 Barcelona, Spain.

of promoter methylation status (Supplementary Fig. 4b)[15]. Indeed, a complete loss of progerin mRNA in HGPS-iPSCs was observed (Fig. 2a). Furthermore, expression of lamin A/C and progerin proteins was practically undetectable (Fig. 2b and Supplementary Fig. 4c).

Because HGPS-iPSCs did not express progerin, we examined whether the nuclear abnormalities observed in HGPS fibroblasts would also be absent at the pluripotent stage. Our results indicate that all of the epigenetic, nuclear lamina and proliferation parameters analysed in HGPS-iPSCs are indistinguishable from BJ-iPSCs (Fig. 2b, c and Supplementary Figs 3 and 7). In addition, the nuclei of HGPS-iPSCs displayed the characteristic wrinkles and lobes observed in hESCs and iPSCs (Supplementary Fig. 8), indicative of a reprogramming of the nuclear envelope components. As the nuclear envelope associates with and regulates heterochromatin[11,16], we next examined genome-wide CpG methylation in HGPS fibroblasts, BJ fibroblasts, HGPS-iPSCs, BJ-iPSCs, and H9 hESCs. Using bisulphite padlock probes and Illumina sequencing, we captured and quantified the methylation level of an average of 95,932 CpG sites within a set of 16,206 well-annotated differentially methylated regions (DMRs)[17] per cell line (Supplementary Table 1). The correlation coefficient of the global methylation levels between the pluripotent lines (BJ-iPSCs, HGPS-iPSCs and H9 hESCs) and the corresponding fibroblasts indicated that the generated pluripotent lines are much more closely related to each other and to hESCs than the two fibroblast lines (Fig. 2d). Interestingly, 586 autosome genes were found to be associated with regions that showed methylation differences between HGPS and BJ fibroblasts (Supplementary Table 2 and Supplementary Fig. 9a). Furthermore, based on DAVID[18,19] analysis, we found that these genes were enriched for 21 Gene Ontology terms, most of which were related to development and transcriptional regulation (Supplementary Fig. 9a). In contrast, methylation differences between HGPS-iPSCs and BJ-iPSCs were only found for 33 autosome genes (Supplementary Table 3), which showed no significant functional enrichment. Therefore, the presence of progerin in HGPS fibroblasts seems to lead

to major epigenomic changes in various pathways. These changes were no longer present in HGPS-iPSCs, coinciding with the downregulation of progerin. Finally, genome-wide mRNA profiling demonstrated that HGPS-iPSCs and BJ-iPSCs are closely related together with H9 hESCs, and different from their parental fibroblasts (Supplementary Fig. 9b–d). These results demonstrate the complete resetting of the nuclear architecture, epigenome and global gene expression in HGPS cells after being reprogrammed to pluripotency.

To test whether the expression of progerin could be re-activated, we first subjected HGPS-iPSCs to in vitro differentiation via EB formation. Progerin mRNA was selectively induced in differentiated HGPS-iPSCs, but not in differentiated BJ-iPSCs (Supplementary Fig. 10a). In contrast, lamin A was upregulated in both differentiated HGPS-iPSCs and differentiated BJ-iPSCs (Supplementary Fig. 10a). This reversible suppression of progerin expression by reprogramming, and subsequent reactivation upon differentiation, provides a unique model system to study human premature ageing pathologies. Progerin is known to accumulate mainly in arterial SMCs of HGPS patients, and vascular SMC degeneration is one of the characteristics of HGPS-associated arteriosclerosis[6,7,20]. In fact, vascular SMC senescence has been involved in the advanced arteriosclerosis of normal populations[7,21,22]. Therefore, we next asked whether SMCs differentiated from HGPS-iPSCs exhibit premature senescence phenotypes. Using a directed differentiation protocol, we obtained SMC populations from HGPS-iPSCs and BJ-iPSCs, the majority of which expressed characteristic SMC markers such as smooth muscle actin (SMA) and calponin (Supplementary Fig. 6a). Immunoblotting and PCR with reverse transcription (RT–PCR) analyses confirmed the expression of progerin in HGPS-iPSC, but not BJ-iPSC-derived SMCs (hereafter referred to as 'HGPS-SMC' and 'BJ-SMC', Supplementary Fig. 6b, c). To model SMC senescence in vitro, the differentiated SMCs were serially passaged in culture. As shown in Fig. 3a–c and Supplementary Fig. 10b, an increasing frequency of misshapen nuclei and a loss of the heterochromatin mark H3K9me3
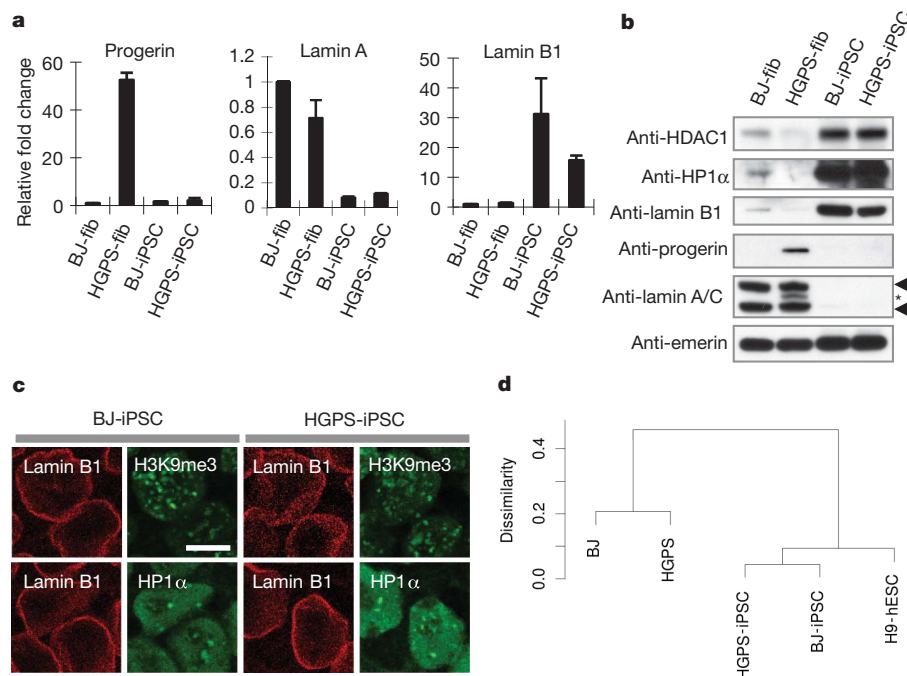

Figure 2 | HGPS-associated nuclear defects are reset in HGPS-iPSCs. a, RT–PCR analysis of progerin, lamin A and lamin B1 in the specific cell lines (n = 3). BJ- and HGPS-fib, BJ and HGPS fibroblasts. **b**, Immunoblotting analysis of the indicated proteins. Emerin was used as loading control. Asterisk denotes progerin (Δ50 lamin A). Arrowheads denote lamin A (top) and lamin C (bottom). **c**, Immunofluorescence analysis performed on BJ-iPSCs and HGPS-iPSCs for detection of the indicated proteins. Scale bar, 10 μm. **d**, Hierarchical clustering of genome-wide DNA methylation profiles.
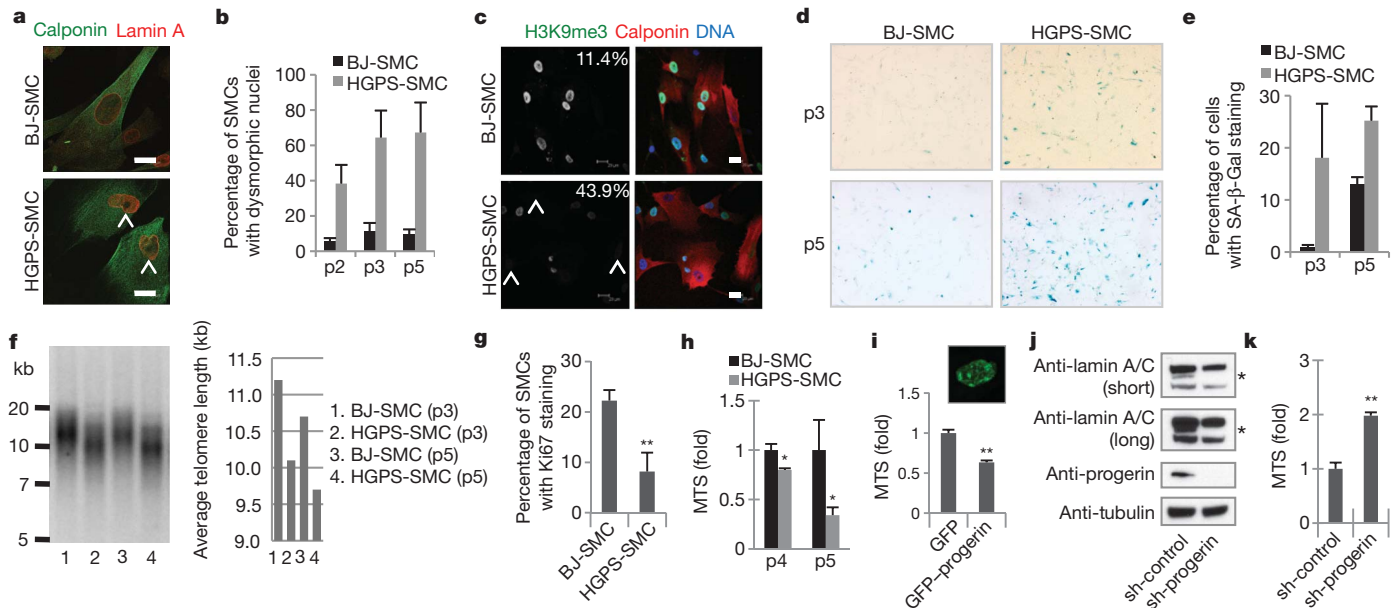
Figure 3 | SMCs expressing progerin show nuclear defects and accelerated senescence. a, Immunostaining of calponin and lamin A in iPSC-derived SMCs (p5). Arrowheads denote dysmorphic nuclei. Scale bar, 20 μm. **b**, Percentage of calponin-positive cells showing dysmorphic nuclei, ($n = 3$, $P < 0.001$). p2, p3 and p5, passage 2, 3 and 5, respectively. **c**, Immunostaining of H3K9me3 and calponin in iPSC-derived SMCs (p5). Nuclei were visualized with Hoechst stain (blue). Arrowheads denote decreased nuclear H3K9me3 (percentage in corner). Scale bar, 20 μm. **d**, **e**, Senescence-associated (SA)-β-Gal staining of iPSC-derived SMCs, $P < 0.05$. **f**, Southern blot analysis of SMCs showing telomere length (left). Quantified average of telomere length (right,

$n = 2$). **g**, Percentage of Ki67-positive cells in iPSC-derived SMCs (calponin-positive, p3), $**P < 0.01$. **h, i**, Cell proliferation analysis of iPSC-derived SMCs ($n = 3$), $*P < 0.05$ (**h**) or primary vascular SMCs (overexpressing GFP or GFP–progerin, $n = 3$), $**P < 0.01$ (**i**). Typical GFP–progerin-positive nucleus showing abnormal morphology (inset). **j**, Immunoblotting of the indicated proteins in shRNA-modified HGPS-iPSCs after 21 days of EB-mediated differentiation. Asterisks denote progerin (Δ50 lamin A). **k**, Cell proliferation analysis of the SMCs derived from shRNA-modified HGPS-iPSCs (p2, $n = 3$), $**P < 0.01$.

were specifically observed in HGPS-SMCs after serial passaging. HGPS-SMCs at later passages (that is, passage 5) showed the typical characteristics of premature senescence, including increased senescence-associated-β-galactosidase (SA-β-Gal) staining (Fig. 3d, e and Supplementary Fig. 10c), reduced telomere length (Fig. 3f), a reduced number of Ki67-positive cells (Fig. 3g and Supplementary Fig. 10d), and compromised cell proliferation (Fig. 3h and Supplementary Fig. 10e). We also found a selective upregulation of senescence-related transcripts in HGPS-SMCs (Supplementary Fig. 10f). To test whether the observed HGPS-related cell phenotypes were specific to SMCs, we differentiated HGPS-iPSCs into fibroblasts and measured progerin-associated parameters. Progerin expression in HGPS-iPSC-derived fibroblasts was detectable as early as passage 5 (Supplementary Fig. 11a). However, we were unable to detect a loss of lamina or heterochromatin markers before passage 10 (Supplementary Fig. 11b, c). Nonetheless, these defects were present specifically in HGPS-iPSC-derived fibroblasts, in contrast to control iPSC-derived fibroblasts analysed at similar passage (data not shown). Thus, even though direct comparison of SMCs and fibroblasts is not possible due to their different culture conditions, our observations demonstrate that mesoderm lineages derived from HGPS-iPSCs display a characteristic HGPS phenotype.

We next investigated whether progerin accumulation is the direct cause of the accelerated cell senescence observed in HGPS-SMCs. To this end, we induced ectopic expression of progerin in human primary vascular SMCs. We found that introduction of progerin in wild-type SMCs resulted in compromised cell proliferation and nuclear defects, as we had observed in HGPS-SMCs (Fig. 3i and Supplementary Fig. 12). As a complementary approach, we transduced HGPS-iPSCs with a lentiviral vector expressing a progerin-specific short hairpin RNA (shRNA)[23]. The modified iPSCs showed normal karyotypes as well as normal expression of lamina/epigenetic and pluripotent markers

(Supplementary Fig. 13a, b). After EB-based differentiation, both the mRNA and protein levels of progerin, but not those of lamin A, were substantially downregulated in the progerin-shRNA 'corrected' HGPS-iPSCs compared to control cells (Fig. 3j and Supplementary Fig. 13c, d). We next differentiated these 'progerin-free' HGPS-iPSCs into SMCs (Supplementary Fig. 14a). A marked improvement in the proliferation capability, as well as a downregulation of senescence-related transcripts, was found in the SMCs differentiated from the corrected HGPS-iPSCs (Fig. 3k and Supplementary Fig. 14b, c). Furthermore, transduction of progerin shRNA into early passage HGPS-iPSC-derived fibroblasts resulted in a clear restoration of nuclear morphology and heterochromatin markers after extended culture (Supplementary Fig. 15). Taken together, these data identify progerin as the key factor underlying the premature senescence phenotypes of HGPS-iPSC-derived cells.

Because phenotypic characteristics of premature ageing were able to be recapitulated by directed differentiation of the HGPS-iPSCs, we next investigated whether this model could serve to identify novel senescence-related markers. By using a sensitive MudPIT proteomic approach[24,25], we identified DNA-dependent protein kinase catalytic subunit (DNAPKcs) as a hitherto unknown binding partner of progerin (Supplementary Table 4). DNAPK holoenzyme, comprising DNAPKcs and its regulatory subunits Ku70/Ku80 (also known as XRCC6 and XRCC5, respectively), is involved in various ageing-related cellular events[26,27], and DNAPKcs or Ku70/Ku80-deficient mice exhibit accelerated ageing[27,28]. To further confirm the association of progerin with DNAPKcs, we performed co-immunoprecipitation experiments. As shown in Fig. 4a, ectopically expressed progerin associated tightly with endogenous DNAPKcs. In contrast, lamin A showed weak interaction with DNAPKcs, whereas both progerin and lamin A exhibited similar binding to lamin B1. Neither progerin nor lamin A co-immunoprecipitated with endogenous WRN protein. Since most of
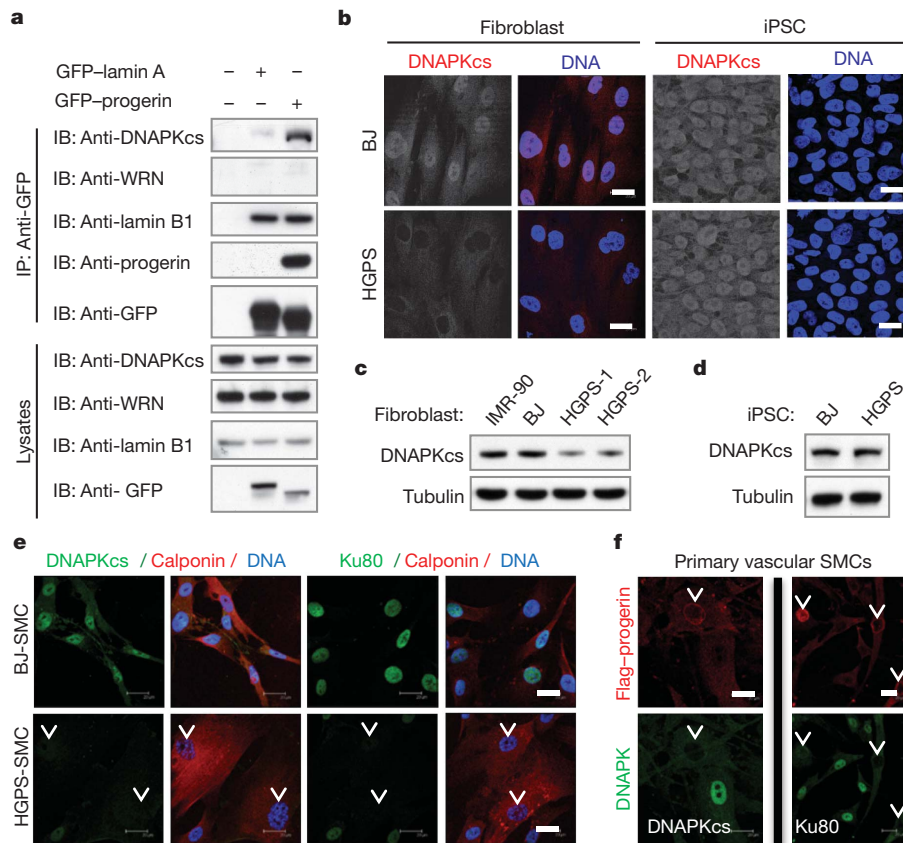
Figure 4 | Decreased expression of DNAPK holoenzyme correlates with premature cell ageing. a, Extracts from BJ fibroblasts expressing GFP, GFP–lamin A, or GFP–progerin, were immunoprecipitated (IP) with a GFP antibody and examined by immunoblotting analysis (IB). **b**, DNAPKcs staining in the indicated cell lines. **c, d**, Immunoblot analysis of DNAPKcs expression. **e, f**, Immunostaining of the indicated proteins in iPSCs-derived SMCs (**e**) or primary vascular SMCs overexpressing progerin (**f**). Arrowheads denote decreased DNAPKcs or Ku80. Nuclei were visualized with Hoechst stain (blue). Scale bar, 20 μm.

the nuclear proteins in complex with lamin A are destabilized in HGPS cells[9,29], we analysed the status of DNAPKcs in primary HGPS fibroblasts. We observed decreased nuclear DNAPKcs in HGPS fibroblasts when compared to normal fibroblasts (Fig. 4b, c). In addition, the regulatory subunits Ku70/Ku80 were also downregulated in HGPS fibroblasts (Supplementary Fig. 16a). Interestingly, we detected a complete restoration of DNAPKcs/Ku70/Ku80 expression in HGPS-iPSCs (Fig. 4b, d and Supplementary Figs 3 and 16b), although a deficiency in the expression of these proteins reappeared after differentiation into SMCs (Fig. 4e). These observations indicate that the downregulation of DNAPKcs in HGPS cells is dependent on the accumulation of progerin in differentiated cells. In fact, ectopic expression of progerin in primary vascular SMCs diminished DNAPKcs/Ku80 protein expression (Fig. 4f and Supplementary Fig. 17). We next tested whether inactivation of DNAPKcs could partially contribute to some of the phenotypes observed in HGPS-SMCs. We found that knockdown of DNAPKcs reduced the proliferation of primary vascular SMCs (Supplementary Fig. 18a). Finally, we extended our results and found that progressive loss of DNAPKcs/Ku70/Ku80 also occurs in fibroblasts isolated from normally ageing individuals (Supplementary Fig. 18b). Overall, our data indicate that deficiency of the DNAPK holoenzyme may constitute a novel marker for premature as well as physiological ageing.

In summary, our results not only highlight the plasticity of the lamina-epigenetics axis, but also point to the fact that the altered structure of the nuclear envelope, as well as the epigenetic modifications that accumulate during physiological ageing[12] or under specific disease conditions[1,11], can be restored to normalcy by reprogramming (Supplementary Fig. 19). The gradual onset and complexity of ageing has impeded progress in understanding the pathogenesis of ageing-related

cardiovascular disorders. Recently, striking similarities between normal ageing-associated and HGPS-associated arteriosclerosis have been reported[6,16]. Indeed, the levels of progerin increase gradually during physiological ageing[6]. Our study provides the first evidence that, in a progerin-dependent manner, HGPS-iPSC-derived SMCs reach senescence-related phenotypes earlier than their normal counterparts. The iPSC-based accelerated ageing model presented here and in ref. 30 may provide an avenue to model and study the pathogenesis of human ageing-related vascular diseases as well as various human laminopathies[1].

METHODS SUMMARY

iPSCs were generated from human fibroblasts with retroviruses encoding OCT4/SOX2/KLF4/c-MYC/GFP, and grown on MEF feeder cells or Matrigel. SMCs were differentiated from iPSCs-derived CD34+ progenitor cells following an OP9-based protocol.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Worman, H. J., Ostlund, C. & Wang, Y. Diseases of the nuclear envelope. *Cold Spring Harb. Perspect. Biol.* **2,** a000760 (2010).
2. Burtner, C. R. & Kennedy, B. K. Progeria syndromes and ageing: what is the connection? *Nature Rev. Mol. Cell Biol.* **11,** 567–578 (2010).
3. Kudlow, B. A., Kennedy, B. K. & Monnat, R. J. Jr. Werner and Hutchinson–Gilford progeria syndromes: mechanistic basis of human progeroid diseases. *Nature Rev. Mol. Cell Biol.* **8,** 394–404 (2007).
4. Merideth, M. A. *et al.* Phenotype and course of Hutchinson–Gilford progeria syndrome. *N. Engl. J. Med.* **358,** 592–604 (2008).

5. Davies, B. S., Fong, L. G., Yang, S. H., Coffinier, C. & Young, S. G. The posttranslational processing of prelamin A and disease. *Annu. Rev. Genomics Hum. Genet.* **10,** 153–174 (2009).

6. Olive, M. *et al.* Cardiovascular pathology in Hutchinson-Gilford progeria: correlation with the vascular pathology of aging. *Arterioscler. Thromb. Vasc. Biol.* doi:10.1161/ATVBAHA.110.209460 (26 August 2010).

7. Ragnauth, C. D. *et al.* Prelamin A acts to accelerate smooth muscle cell senescence and is a novel biomarker of human vascular aging. *Circulation* **121,** 2200–2210 (2010).

8. Varga, R. *et al.* Progressive vascular smooth muscle cell defects in a mouse model of Hutchinson–Gilford progeria syndrome. *Proc. Natl Acad. Sci. USA* **103,** 3250–3255 (2006).

9. Pegoraro, G. *et al.* Ageing-related chromatin defects through loss of the NURD complex. *Nature Cell Biol.* **11,** 1261–1267 (2009).

10. Scaffidi, P. & Misteli, T. Reversal of the cellular phenotype in the premature aging disease Hutchinson-Gilford progeria syndrome. *Nature Med.* **11,** 440–445 (2005).

11. Dechat, T. *et al.* Nuclear lamins: major factors in the structural organization and function of the nucleus and chromatin. *Genes Dev.* **22,** 832–853 (2008).

12. Scaffidi, P. & Misteli, T. Lamin A-dependent nuclear defects in human aging. *Science* **312,** 1059–1063 (2006).

13. McClintock, D. *et al.* The mutant form of lamin A that causes Hutchinson-Gilford progeria is a biomarker of cellular aging in human skin. *PLoS ONE* **2,** e1269 (2007).

14. Constantinescu, D., Gray, H. L., Sammak, P. J., Schatten, G. P. & Csoka, A. B. Lamin A/C expression is a marker of mouse and human embryonic stem cell differentiation. *Stem Cells* **24,** 177–185 (2006).

15. Freberg, C. T., Dahl, J. A., Timoskainen, S. & Collas, P. Epigenetic reprogramming of OCT4 and NANOG regulatory regions by embryonal carcinoma cell extract. *Mol. Biol. Cell* **18,** 1543–1553 (2007).

16. Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* **38,** 603–613 (2010).

17. Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genet.* **41,** 178–186 (2009).

18. Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4,** 44–57 (2009).

19. Dennis, G. *et al.* DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* **4,** R60 (2003).

20. McClintock, D., Gordon, L. B. & Djabali, K. Hutchinson–Gilford progeria mutant lamin A primarily targets human vascular cells as detected by an anti-Lamin A G608G antibody. *Proc. Natl Acad. Sci. USA* **103,** 2154–2159 (2006).

21. Gorenne, I., Kavurma, M., Scott, S. & Bennett, M. Vascular smooth muscle cell senescence in atherosclerosis. *Cardiovasc. Res.* **72,** 9–17 (2006).

22. Minamino, T. & Komuro, I. Vascular cell senescence: contribution to atherosclerosis. *Circ. Res.* **100,** 15–26 (2007).

23. Huang, S. *et al.* Correction of cellular phenotypes of Hutchinson-Gilford progeria cells by RNA interference. *Hum. Genet.* **118,** 444–450 (2005).

24. Liu, G. H. *et al.* Regulation of myoblast differentiation by the nuclear envelope protein NET39. *Mol. Cell. Biol.* **29,** 5800–5812 (2009).

25. Washburn, M. P., Wolters, D. & Yates, J. R., III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.* **19,** 242–247 (2001).

26. Ruis, B. L., Fattah, K. R. & Hendrickson, E. A. The catalytic subunit of DNA-dependent protein kinase regulates proliferation, telomere length, and genomic stability in human somatic cells. *Mol. Cell. Biol.* **28,** 6182–6195 (2008).

27. Li, H., Vogel, H., Holcomb, V. B., Gu, Y. & Hasty, P. Deletion of Ku70, Ku80, or both causes early aging without substantially increased cancer. *Mol. Cell. Biol.* **27,** 8205–8214 (2007).

28. Espejel, S. *et al.* Shorter telomeres, accelerated ageing and increased lymphoma in DNA-PKcs-deficient mice. *EMBO Rep.* **5,** 503–509 (2004).

29. Han, X. *et al.* Tethering by lamin A stabilizes and targets the ING1 tumour suppressor. *Nature Cell Biol.* **10,** 1333–1340 (2008).

30. Zhang, J. *et al.* A human iPSC model of Hutchinson Gilford progeria reveals vascular smooth muscle and mesenchymal stem cell defects. *Cell Stem Cell* **8,** 31–45 (2011).

Author Contributions G.-H.L. and J.C.I.B. conceived the experiments; G.-H.L., B.Z.B., S.R., D.D., J.Q., S.-L.Y., A.D.P., K.S., L.K., C.W., J.T. and H.L.F. performed the experiments and analysed the data; S.B., I.S.-M., K.Z., J.Y. and J.C.I.B. analysed the data; G.-H.L., S.R., B.Z.B., A.D.P., K.Z. and J.C.I.B. wrote the manuscript.

METHODS

Cell culture. H9 hESCs (WiCell Research) and iPSCs were maintained on a layer of mitotically inactivated mouse embryonic fibroblasts (MEFs) in hESC medium: DMEM/F12 (Invitrogen) supplemented with 0.1 mM non-essential amino acids (Invitrogen), 1 mM GlutaMAX (Invitrogen), 20% Knockout Serum Replacement (Invitrogen), 55 μM β-mercaptoethanol (Invitrogen) and 10 ng ml^{-1} bFGF (Joint Protein Central). hESCs and iPSCs were also cultured in Matrigel (BD Biosciences) with mTeSR medium (Stem Cell Technologies). Human HGPS fibroblasts AG01972, AG11498, AG06297, and normal fibroblasts GM00038 (9 years), AG05247 (87 years), and AG09602 (92 years) were purchased from Coriell Cell Repository. BJ normal human fibroblasts (CRL-2522) were purchased from ATCC. All human fibroblasts were cultured at 37 °C in DMEM containing GlutaMAX, non-essential amino acids, sodium pyruvate, and 15% fetal bovine serum (FBS). Human aortic smooth muscle cells were purchased from Lonza and maintained in SmGM-2 medium (Lonza, Cat. # CC-3182).

Reagents. Antibodies were obtained from the following sources. Abcam: anti-NANOG (ab21624), anti-H3K9me3 (ab8898), anti-progerin (ab66587), anti-emerin (ab14208); anti-Ku70 (ab2172); Santa Cruz Biotechnology: anti-Oct-3/4 (sc-5279), anti-SOX2 (sc-17320), anti-HDAC1 (sc-7872), anti-DNAPKcs (sc-9051), anti-lamin A/C (sc-6215), anti-lamin A/C (sc-7293), anti-lamin B1 (sc-6217); Cell Signaling: anti-HP1α (2616); anti-Ku80 (2753); R&D systems: anti-Foxa2 (AF2400); Millipore: anti-TRA-1-60 (MAB4360) ; Sigma: anti-β-Tubulin III (T2200), anti-SMA(A5228), anti-Flag (M2), and anti-tubulin (T5168); Dako: anti-calponin (clone CALP); anti-endoglin (clone SN6h); BD Transduction Laboratories: anti-LAP2β (611000); MBL: agarose-conjugated anti-GFP.

Plasmids. The pMXs vector containing the human cDNAs for *OCT4*, *SOX2*, *KLF4* and *c-MYC* were purchased from Addgene (17217, 17218, 17219 and 17220, respectively). pBABE-puro-GFP-progerin and pBABE-puro-GFP-wt-lamin A were purchased from Addgene (17663 and 17662, respectively). Flag-progerin lentiviral vector was kindly provided by L. Comai[31]. For the generation of the shRNA expression vectors against progerin and DNAPKcs, corresponding oligo-nucleotides (see Supplementary Table 5) were cloned into a MluI–ClaI-cleaved pLVTHM plasmid (Addgene, 12247). All the constructs generated were subjected to DNA sequencing to confirm accurate shRNA target sequence.

Retrovirus and lentivirus production. For retrovirus production, 293T cells were transfected with the pMXs vectors carrying *OCT4*, *SOX2*, *c-MYC*, *KLF*4 or *GFP* cDNAs, together with the packaging plasmids (pCMV-gag-pol-PA and pCMV-VSVg, provided by G. Pao, The Salk Institute) using Lipofectamine 2000 (Invitrogen). Retroviruses were collected 36–48 h after transfection, and filtered through a 0.45-μM filter. Lentiviruses were generated by co-transfecting the pLVTHM vector together with the packaging plasmids (psPAX2 and pMD2.G, from Addgene, 12260 and 12259 respectively) into 293T cells using Lipofectamine 2000 (Invitrogen). Lentiviruses were collected 36 h after transfection and concentrated by ultracentrifugation.

iPSCs generation. For the generation of human iPSCs, human fibroblasts were seeded in a 6-well plate and spin-infected with a mix of high-quality retroviruses encoding OCT4, SOX2, KLF4, c-MYC and GFP in the presence of 4 μg ml^{-1} polybrene. Three infections on consecutive days were performed. Six days after the first infection, fibroblasts were gently individualized with TrypLE (Invitrogen) and seeded onto fresh MEFs in the fibroblast culture medium. After 24 h, the medium was switched to hESC medium, and changed every 1–2 days depending on cell density. To establish the iPSC lines, colonies were manually picked and transferred onto MEF feeder cells for several passages before being transferred to Matrigel/mTeSR conditions.

Lentiviral infection of iPSCs. HGPS-iPSC#4 cell line cultured on Matrigel was treated with 10 μM ROCK inhibitor Y-27632 for 1 h and then individualized with TrypLE. Cells were infected in suspension with either the concentrated lentivirus pLVTHM or pLVTHM-shRNA-progerin in the presence of ROCK inhibitor and polybrene (4 μg ml^{-1}) for 1 h. Cells were centrifuged to remove the lentivirus and seeded back on fresh feeder MEFs in hESC media containing ROCK inhibitor. After growing for a few days, small colonies were manually passaged as a pool of colonies onto fresh MEFs to establish new iPSC lines. GFP expression was used as an indicator to determine successful integration of the lentiviruses.

Cell differentiation. For embryoid bodies (EBs) based differentiation, the iPSC colonies growing on MEFs were detached with dispase treatment, resuspended in DMEM/F12 medium supplemented with 0.1 mM non-essential amino acids, 0.5 mM L-glutamine, 10% FBS (Atlanta Biologicals), and 55 μM β-mercaptoethanol and cultured in low attachment 6-well plates for 4 days. The EBs were then plated on gelatin-coated plates and maintained for another 10–17 days. Differentiation of iPSCs into fibroblasts was performed as previously described[32]. Directed differentiation towards smooth muscle cells (SMCs) was performed essentially as previously described[33] with slight modifications. Irritated OP9 cells were plated at 1×10^5 cells per well onto gelatinized 6-well plates in OP9 growth medium. After the formation

of confluent iPS cell cultures for 4 and 5 days, undifferentiated iPS cells were harvested by treatment with 1 mg ml^{-1} dispase and dispersed by scraping to maintain the cells in small clumps. Concurrently, iPS cultures growing under the same conditions were used to obtain single cell suspension for counting. The iPSCs were added to OP9 cultures at a density of 3×10^5 per 2 ml per well of a 6-well plate in half TeSR medium and half hESC media. iPSCs were allowed to recover for 1–2 days in hESC media. At day 0 of differentiation, the media was changed to Knockout DMEM supplemented with 10% FBS (HyClone), 10 mM β-mercaptoethanol, 1 mM L-glutamine, and 100 mM nonessential amino acids. The iPSC/OP9 co-cultures were incubated for up to 10 days at 37 °C in 5% CO2 conditions with medium change every other day. After 10 days of differentiation, the co-culture cells were harvested with TrypLE (Invitrogen) for single-cell suspension and labelled with CD34 microbeads kit (Miltenyi Biotech, Cat. #130-046-702). Following the manufacturer's protocol, cells were passed through MS separation column attached to a Midi-MACS separation unit (Miltenyi Biotech) to obtain a magnet-retained fraction of purified CD34$^+$ cells. Isolated CD34$^+$ cells were then plated in smooth muscle cell media (SmGM-2 BulletKit, Lonza, Cat. # CC-3182) and maintained at 37 °C in 5% CO$_2$ conditions with medium change every 2–3 days[34]. SMCs were passaged using TrypLE (diluted 1:4) for 3 min at 37 °C. To analyse early onset of senescence, cells were passaged at a ratio of 1:3 (~6,000–7,500 cells per cm^2) only when the cells reached confluence. To calculate population doublings, SMCs seeded at 3,500 cells per cm^2 were passaged once culture reached 85–90% confluence. Cell growth was measured at every passage by calculation of accumulated population doublings using the formula (log$H$ − log$S$)/log2.0 where H = number of cells harvested, S = number of cells seeded on the first day of each passage.

Protein and mRNA analysis. Cells were lysed and subjected to immunoblotting analysis according to the previously described method[35]. Total RNA was extracted using TRIzol (Invitrogen) followed by cDNA synthesis using High capability RNA-to-cDNA Mater Mix (Invitrogen). Quantitative RT–PCR was performed using SYBR Green PCR Master Mix (Applied Biosystems). Primer sequences are listed in Supplementary Table 5.

Immunofluorescence microscopy. Cells were fixed with 4% formaldehyde in PBS at room temperature for 20–30 min. After fixation, cells were treated with 0.4% Triton X-100 in PBS for 5 min at room temperature. After blocked with 10% FBS in PBS for 30 min, cells were incubated at room temperature for 1 h or at 4 °C overnight with the primary antibody, followed by washing in PBS and incubation at room temperature for 1 h with the corresponding secondary antibody. Nuclei were stained with Hoechst 33342 (Invitrogen). Quantitative microscopy measurements were carried out as described previously[9]. Error bars represent standard deviations.

Immunohistochemical detection of NANOG. Cells were fixed with 4% form-aldehyde in PBS at room temperature for 30 min, and permeabilized with 0.4% Triton-X100 in PBS for 10 min. Then the cells were incubated overnight with rabbit anti-human NANOG antibody in 1% BSA/PBS, followed by incubation with a secondary biotin-conjugated anti-rabbit antibody for 2 h. Finally, cells were incubated with streptavidin-HRP for 1 h (Vector), and NANOG-positive cells were visualized with a DAB substrate kit (Vector).

Teratoma analysis. To test pluripotency *in vivo*, NOD-SCID IL2Rgamma^{null} mice (Jackson laboratories) were injected with the indicated iPSC lines and teratoma formation assessed. Briefly, ~10^6 iPSCs in ~50 μl of hESC medium were injected into the testis or kidney capsule of anesthetized mice. Mice were monitored for teratoma formation and euthanized ~6–12 weeks after injection. Teratomas were harvested, processed and analysed by haematoxylin–eosin staining and immuno-staining. All animal experiments were performed with approval of The Salk Institute Institutional Animal Care and Use Committee (IACUC).

Mutation validation. Primer sequences to amplify exon 11 of the *LMNA* gene are listed in Supplementary Table 5. PCR (final volume 50 μl) reactions using 3 ng genomic DNA templates, 100 nM of the forward and reverse primers with 25 μl Taq 2× Master Mix (NEB) was performed at 94 °C for 2 min, 34 cycles of 94 °C 30 s, 55.5 °C for 40 s, and 72 °C for 40 s, and finally 72 °C for 3 min. Products were purified with 0.9× volume of AMPure beads (Agencourt). Amplicons were sequenced by capillary Sanger sequencing (Genewiz). Results were visualized using an ABI Sequence Scanner.

Genome-wide DNA methylation analysis. Genomic DNA was extracted using ALLPrep DNA/RNA Mini kit (Qiagen). Bisulphite conversion and capture reac-tion was carried out on each sample (genomic DNA of fibroblasts, iPSCs, or hESCs). The detailed protocol for genomic DNA methylation has been described previously[36], and the detailed information for DNA methylation is presented in Supplementary Table 1.

Bisulphite sequencing of OCT4 and *LMNA* promoters. Bisulphite conversion was carried out using 2 μg of purified genomic DNA using the Zymo EZ-DNA Methylation Gold Kit (Zymo Research) following the manufacturer's instructions. PCR was set up using previously published primers[15]. Cycling was terminated at 35

cycles. PCR products were purified using 2% Size-Select E-gel (Invitrogen) and reamplified for 10 cycles using Phusion HF enzyme (NEB). PCR products were cloned using Zero-blunt PCR Cloning kit (Invitrogen) and heat transfected to TOP10 *Escherichia coli* competent cells (Invitrogen). Individual colonies were selected and sent for single pass sequencing.

DNA microarray and bioinformatics analysis. The GeneChip microarray processing was performed by the Functional Genomica Core in the Institute for Research in Biomedicine (Barcelona, Spain) according to the manufacturer's protocols (Affymetrix). The amplification and labelling were processed as indicated in Nugen protocol with 25 ng starting RNA. For each sample, 3.75 μg ssDNA were labelled and hybridized to the Affymetrix HG-U133 Plus 2.0 chips. Expression signals were scanned on an Affymetrix GeneChip Scanner (7G upgrade). The data extraction was done by the Affymetrix GCOS software v.1.4.

The statistical analysis of the data was performed using ArrayStar 3. Briefly, raw CEL files were imported together with gene annotation from NetAffix (from 11/13/2009) and after checking for top replication quality for each of the five pairs of samples ($R^2 > 0.99$), data was summarized at the gene level (20,765 genes) and the median was used for each gene and sample type. As both H9 hESCs and HGPS-iPSCs originate from female samples, and in order to remove any possible bias introduced by the X and Y chromosome-coded genes, we performed the same analysis with only autosome genes (19,884 genes). The result of the hierarchical clustering is very similar to the one using all genes and is shown in Supplementary Fig. 9b. In addition, a principal component analysis was performed on RMA-normalized probeset intensity values for autosomes using the prcomp function in R (http://www.r-project.org/) (the same figure including all genes gave highly similar results, data not shown). A figure illustrating the two first principal components is shown in Supplementary Fig. 9c. Differences between some of the samples is shown using scatter plot of RMA-normalized intensity values in Supplementary Fig. 9d.

Multidimensional protein identification technology (MudPIT) analysis of progerin-associated proteins. The immunoprecipitation for MudPIT assay was performed as previously described[24,25]. In brief, HEK293T cells were transfected with GFP–progerin or GFP and maintained in culture for 48 h. After cells were lysed, the GFP–progerin, GFP, and their associated proteins were immunoadsorbed to anti-GFP agarose. The immunoprecipitates were then eluted with 8 M urea in 100 mM Tris, pH 8.5. The samples were reduced by adding 0.3 μl of 1 M TCEP (for a final concentration of 5 mM TCEP) and incubated at room temperature. To alkylate, 1.2 μl of iodoacetamide (10 mM final concentration) was added and the samples were subsequently incubated at room temperature in the dark for 15 min. The addition of 180 μl of 100 mM Tris pH 8.5 diluted the solutions to 2 M urea. Calcium chloride (100 mM) was then added (2.4 μl) for a final concentration of 1 mM CaCl₂. Trypsin (0.5 μg μl⁻¹) was added in the amount of 7.0 μl. The resulting mixtures were then shaken for 18 h and incubated in the dark at 37 °C. To neutralize 13.5 μl of formic acid (90%) was added for a final concentration of 5% formic acid. The tubes were centrifuged for 30 min at 2 °C in a table-top centrifuge. Upon completion of the digestion, the proteins were pressure-loaded onto a fused silica capillary desalting column containing 3 cm of 5-μm strong cation exchange (SCX) followed by 3 cm of 5-μm C18 (reverse phase or RP material) packed into an undeactivated 250-μm i.d capillary. Using 1.5 ml of buffer A (95% water, 5% acetonitrile and 0.1% formic acid) the desalting columns were washed overnight. Following the desalting process, a 100-μm internal diameter capillary consisting of a 10-μm laser pulled tip packed with 10 cm 3-μm Aqua C18 material (Phenomenex) was attached to the filter union (desalting column–filter union–analytical column) and the entire split-column (desalting column–filter union–analytical column) was placed in line with an Agilent 1100 quaternary HPLC (Palo Alto, CA) and analysed using a modified 6-step separation, described previously[25]. The buffer solutions used were 5% acetonitrile/0.1% formic acid (buffer A), 80% acetonitrile/0.1% formic acid (buffer B), and 500 mM ammonium acetate/5% acetonitrile/0.1% formic acid (buffer C). Step 1 consisted of a 90 min gradient from 0–100% buffer B. Steps 2–5 had the following profile: 3 min of 100% buffer A, 2 min of X% buffer C, a 10 min gradient from 0–15% buffer B, and a 97 min gradient from 15–45% buffer B. The 2 min buffer C percentages (X) were 20, 40, 60, 80% respectively for the 6-step analysis. In the final step, the gradient contained: 3 min of 100% buffer A, 20 min of 100% buffer C, a 10 min gradient from 0–15% buffer B, and a 107 min gradient from 15–70% buffer B. As peptides eluted from the microcapillary column, they were electrosprayed directly into an LTQ 2-dimensional ion trap mass spectrometer (ThermoFinnigan) with the application of a distal 2.4 kV spray voltage. A cycle of one full-scan mass spectrum (400–1,400 *m/z*) followed by 8 data-dependent MS/MS spectra at a 35% normalized collision energy was repeated continuously throughout each step of the multidimensional separation. Application of mass

spectrometer scan functions and HPLC solvent gradients were controlled by the Xcalibur data system.

As each step was executed, its spectra were recorded to a RAW file. This data was then converted into .ms2 format through the use of RawXtract (Version 1.9). From the .ms2 files, poor quality spectra were removed from the data set using an automated spectral quality assessment algorithm[37]. MS/MS spectra remaining after filtering were searched with the SEQUEST algorithm[38] against the NCBI RefSeq Human (04-23-2010) protein database concatenated to a decoy database in which the sequence for each entry in the original database was reversed[25,39]. All searches were parallelized and performed on a Beowulf computer cluster consisting of 100 1.2 GHz Athlon CPUs[40]. No enzyme specificity was considered for any search. SEQUEST results were assembled and filtered using the DTASelect (version 2.0) program. DTASelect 2.0 uses a linear discriminant analysis to dynamically set XCorr and DeltaCN thresholds for the entire data set to achieve a user-specified false positive rate. The false positive rates are estimated by the program from the number and quality of spectral matches to the decoy database. The hits detected uniquely in the GFP–progerin sample but not in GFP sample represent proteins that are specifically associated with progerin, by either direct or indirect interactions.

Co-immunoprecipitation. BJ human fibroblasts were transduced with retrovirus encoding GFP–progerin, GFP–lamin A or GFP, and maintained in culture for 72 h. For immunoprecipitation, cells were lysed in ice-cold lysis buffer (250 mM NaCl, 0.5% Triton X-100, 50 mM Tris, pH 7.5, 1 mM EGTA, 1 mM EDTA, 10% glycerol, and complete protease inhibitor cocktail (Roche Diagnostics)). Samples were briefly sonicated and immunoprecipitated by incubating with anti-GFP agarose. The immunoprecipitates were washed extensively in lysis buffer, eluted in SDS sample buffer, and subjected to immunoblotting.

Senescence-associated beta-galactosidase (SA-βgal) assay. SA-βgal assay were performed based on previously described methods[41].

Measurement of telomere length. Genomic DNA was isolated from 1×10^6 cells. The telomere-specific oligonucleotide probe (5′-TTAGGGTTAGGGTTAGGG TTAGGG-3′; ValueGene) was end-labelled using γ-³²P-ATP (MP Biomedicals) and T4 polynucleotide kinase (NEB). Two μg of genomic DNA for each sample was digested with AluI (NEB) and MboI (NEB) and subjected to Southern analysis with the telomere-specific probe. Mean telomere length was calculated from $\sum OD_i/(\sum OD_i/\sum MW_i)$. OD_i and MW_i are optical density and molecular weight at a given position i, respectively.

Cell proliferation assay. Cell proliferation was determined with CellTiter 96 AQ_ueous One Solution Cell Proliferation Assay (MTS (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium)), according to the protocol provided by the manufacturer (Promega).

Statistical analysis. Results are presented as mean ± s.d. Comparisons were performed with student's *t*-test or one-way anova. $P < 0.05$ was defined as statistically significant.

31. Candelario, J., Borrego, S., Reddy, S. & Comai, L. Accumulation of distinct prelamin A variants in human diploid fibroblasts differentially affects cell homeostasis. *Exp. Cell Res.* **317,** 319–329 (2011).
32. Hockemeyer, D. *et al.* A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. *Cell Stem Cell* **3,** 346–353 (2008).
33. Vodyanik, M. A., Bork, J. A., Thomson, J. A. & Slukvin, I. I. Human embryonic stem cell-derived CD34+ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential. *Blood* **105,** 617–626 (2005).
34. Lu, S. J., Ivanova, Y., Feng, Q., Luo, C. & Lanza, R. Hemangioblasts from human embryonic stem cells generate multilayered blood vessels with functional smooth muscle cells. *Regen. Med.* **4,** 37–47 (2009).
35. Kawamura, T. *et al.* Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature* **460,** 1140–1144 (2009).
36. Deng, J. *et al.* Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nature Biotechnol.* **27,** 353–360 (2009).
37. Bern, M., Goldberg, D., McDonald, W. H. & Yates, J. R., III. Automatic quality assessment of peptide tandem mass spectra. *Bioinformatics* **20** (Suppl 1), i49–i54 (2004).
38. Eng, J., McCormack, A. & Yates, J. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5,** 976–989 (1994).
39. Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J. & Gygi, S. P. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J. Proteome Res.* **2,** 43–50 (2003).
40. Sadygov, R. G. *et al.* Code developments to improve the efficiency of automated MS/MS spectra interpretation. *J. Proteome Res.* **1,** 211–215 (2002).
41. Debacq-Chainiaux, F., Erusalimsky, J. D., Campisi, J. & Toussaint, O. Protocols to detect senescence-associated beta-galactosidase (SA-βgal) activity, a biomarker of senescent cells in culture and *in vivo*. *Nature Protocols* **4,** 1798–1806 (2009).

LETTER

Amyloid–binding compounds maintain protein homeostasis during ageing and extend lifespan

Silvestre Alavez[1], Maithili C. Vantipalli[1], David J. S. Zucker[1,2], Ida M. Klang[1,3] & Gordon J. Lithgow[1]

Genetic studies indicate that protein homeostasis is a major contributor to metazoan longevity[1]. Collapse of protein homeostasis results in protein misfolding cascades and the accumulation of insoluble protein fibrils and aggregates, such as amyloids[2]. A group of small molecules, traditionally used in histopathology to stain amyloid in tissues, bind protein fibrils and slow aggregation in vitro and in cell culture[3,4]. We proposed that treating animals with such compounds would promote protein homeostasis in vivo and increase longevity. Here we show that exposure of adult Caenorhabditis elegans to the amyloid-binding dye Thioflavin T (ThT) resulted in a profoundly extended lifespan and slowed ageing. ThT also suppressed pathological features of mutant metastable proteins and human β-amyloid-associated toxicity. These beneficial effects of ThT depend on the protein homeostasis network regulator heat shock factor 1 (HSF-1), the stress resistance and longevity transcription factor SKN-1, molecular chaperones, autophagy and proteosomal functions. Our results demonstrate that pharmacological maintenance of the protein homeostatic network has a profound impact on ageing rates, prompting the development of novel therapeutic interventions against ageing and age-related diseases.

The longevity of the nematode Caenorhabditis elegans is influenced by hundreds of genes including an insulin-like signalling pathway (ILS) that regulates the activities of the transcription factors FOXO-like DAF-16 (ref. 5) and Nrf2-like SKN-1 (ref. 6). Together with the stress response transcription factor HSF-1, DAF-16 also regulates protein homeostasis and influences lifespan[7–9], indicating that chemical modulation of protein homeostasis might slow ageing. We reasoned that compounds that have protein-fibril- and protein-aggregate-binding properties may affect age-related changes to protein homeostasis and tested a series of amyloid-binding proteins for lifespan effects. We found that exposing sterilized wild-type (N2) nematodes to the fibril-binding flavonoid ThT (4-(3,6-dimethyl-1,3-benzothiazol-3-ium-2-yl)-N,N-dimethylaniline chloride)[10] at either 50 or 100 μM throughout adult life leads to an increase in median (60%) and maximal lifespan (43–78%; Fig. 1a, b, Supplementary Fig. 2 and Supplementary Table 1). The compound reduced age-specific mortality at all ages ($P < 0.001$, Fig. 1c) and slowed age-related decline in spontaneous movement (Fig. 1d), indicating improved health throughout adulthood. At higher doses (500 μM) ThT is toxic and shortens lifespan (Fig. 1a, b). Other compounds with protein-aggregate-binding properties, including curcumin and rifampicin, increased lifespan to a lesser extent (up to 45%) (Supplementary Figs 3, 4). When ThT and curcumin treatments were combined, we did not observe additive effects on lifespan (Supplementary Fig. 5).

We then tested several compounds with similar structural features to ThT, but with different pharmacological properties: 2-(2-hydroxyphenyl)-benzoxazole (HBX), 2-(2-hydroxyphenyl) benzothiazole (HBT) and 2-(2-aminophenyl)-1H-benzimidazole (BM)[11] (Supplementary Fig. 6). These compounds also extended the lifespan of adult nematodes (up to 40%) but at concentrations significantly lower than ThT (Fig. 1e–g),
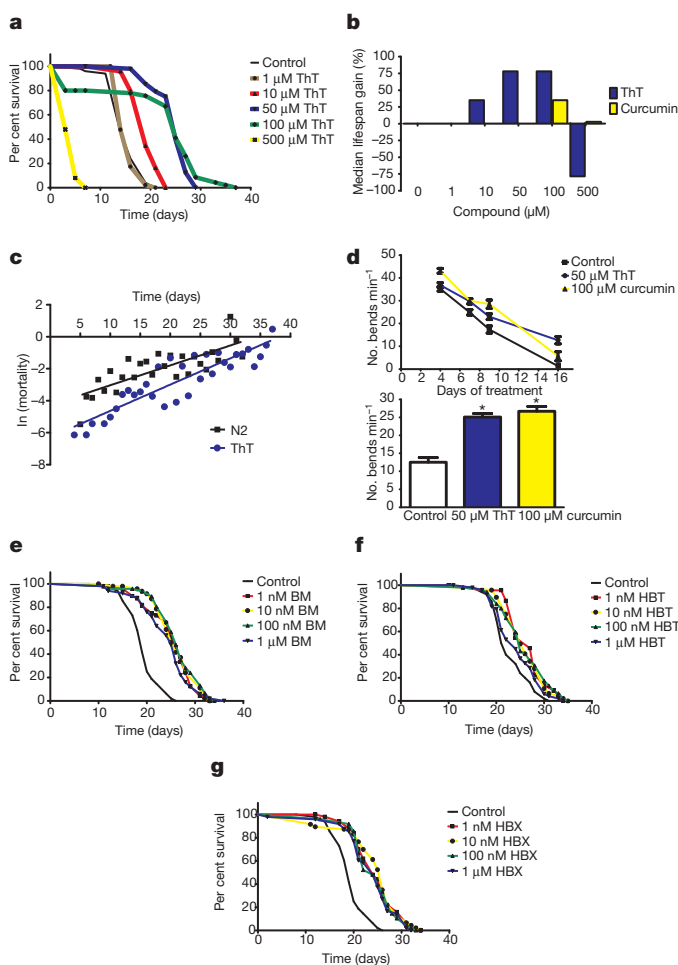

Figure 1 | Amyloid-binding compounds extend C. elegans lifespan.
a, Dose–response Kaplan–Meier survival curves of synchronously ageing hermaphrodite wild-type (N2) populations exposed to 0 μM (control) to 500 μM ThT at 20 °C. **b,** Per cent change in median lifespan of N2 populations cultured on 0–500 μM ThT and curcumin. **c,** ln-linear plot of age-specific mortality rate with age for control and 50 μM ThT-treated C. elegans. **d,** Effect of 50 μM ThT and 100 μM curcumin on motility of N2 worms evaluated as the mean number of body bends in a 20-s period in 15 individual worms throughout life (upper panel) and after 12 days of treatment (lower panel) with ThT and curcumin. Data are presented as bends min^{-1} and represent the average of three independent experiments. *$P < 0.0001$. **e–g,** Dose–response Kaplan–Meier survival curves of synchronously ageing hermaphrodite N2 populations exposed to 0 μM (control) to 1 μM of BM (**e**), HBT (**f**) and HBX (**g**) at 20 °C. Plots are representative of three independent experiments.

[1]Buck Institute for Research on Aging, 8001 Redwood Blvd, Novato, California 94945, USA. [2]Department of Natural Sciences and Mathematics, Dominican University of California, San Rafael, California 94901, USA. [3]Karolinska Institute, Center for Biosciences at NOVUM, Department of Biosciences and Nutrition, Hälsovägen 7, S-141 83 Huddinge, Sweden.

indicating that the bioavailability and/or pharmacological properties of ThT-like compounds influence lifespan.

To test the effects of ThT on protein homeostasis we exploited two *C. elegans* models of human proteotoxic disease: the strain CL4176 (*dvIs27[myo-3::Aβ₃₋₄₂ let 3′UTR(pAF29); pRF4 (rol-6(su1006))])*[12], which expresses an aggregating amyloid-β(3–42) peptide (Aβ(3–42)) in muscle tissue[13] and AM140 (*rmIs132[P(unc-54) Q35::YFP]*), which expresses a polyglutamine (polyQ) protein. Amyloid-β aggregates are associated with lesions in Alzheimer's disease, and polyQ aggregation is a feature in several neurological conditions[14]. When raised at 25 °C, nematodes expressing these proteins in muscle accumulate protein aggregates and become paralysed. We found that 50 μM ThT and 100 μM curcumin decreased the proportion of paralysed worms (Fig. 2a, b). By immunohistochemistry we found that ThT reduced Aβ(3–42) aggregation *in vivo* and preserved muscle integrity in CL4176 (Fig. 2e). We also found that ThT rescued Aβ(3–42) aggregation-induced paralysis even when nematodes were treated 18 h after the induction of aggregate formation, indicating that ThT can ameliorate detrimental effects during the development of the aggregate-related pathology (Supplementary Fig. 7).

If amyloid-binding compounds extended lifespan through improved protein homeostasis, then we expected that they would influence not only heterologous disease-related models but also nematode proteins. We tested ThT and curcumin on mutant worms that express metastable proteins previously exploited as indicators of the protein homeostatic network capacity[15]. Strains carrying mutations in the genes *unc-52* (HE250 [*unc-52(e669su250)II*]) and *unc-54* (CB1157 [*unc-54(e1157)I*]) produce temperature-sensitive muscle proteins UNC-52 (perlecan) and UNC-54 (myosin class II heavy chain), respectively, that exhibit altered structure and cause paralysis at 25 °C[15,16,17]. We found that ThT suppressed paralysis of these mutants (Fig. 2c), prevented the disruption of the muscle sarcomeres (Supplementary Fig. 8) and restored perlecan organization (Fig. 2d). We extended these observations to other temperature-sensitive missense protein-folding mutations expressed in the neuromuscular junction and in the nervous system[18]. We found that ThT suppressed ethanol sensitivity in a strain carrying the *gas-1(fc21)* mutation in a subunit of mitochondrial complex I and levamisole resistance in a strain carrying *unc-63(x26)*, an α-subunit of the nicotinic acetylcholine receptor (Supplementary Fig. 9), indicating that ThT could act in a variety of tissues including the nervous system.

Because certain forms of dietary restriction suppress protein aggregation and increase lifespan, we asked whether ThT acts as a dietary restriction mimetic. We observed that ThT produces a small decrease in pharyngeal pumping rate (~15%) after 3 days of treatment, which could slightly decrease food intake. No difference was detected after 6 days of ThT treatment (Supplementary Fig. 16a). It is very unlikely that this small difference could promote the major ThT-mediated lifespan extension we observe. ThT also increased the lifespan of a strain carrying the *eat-2(ad1116)* mutation (Fig. 4c) that causes a major defect in pharyngeal pumping, thereby inducing a dietary restriction lifespan extension[19]. Dilution of the bacterial food source also leads to lifespan extension by dietary restriction[20]. ThT at 50 μM was detrimental to lifespan in this dietary restriction model but 1 and 10 μM ThT increased lifespan by 24% (Supplementary Fig. 16b). As ThT increases lifespan in both genetic and nutrient-based models of dietary restriction, ThT-induced lifespan extension is at least in part independent from dietary restriction.

We then considered whether ThT was interacting more directly with homeostatic mechanisms. We exploited ThT fluorescence to visualize the compound *in vivo* and observed a variable co-distribution of ThT with Aβ(3–42) aggregates detected by immunolocalization, indicating a direct interaction between ThT protein misfolding cascades (Supplementary Fig. 10). Consequently, we tested for the presence of amino acid sequence-independent oligomers of protein or peptides prone to aggregation. We found that protein detected by an
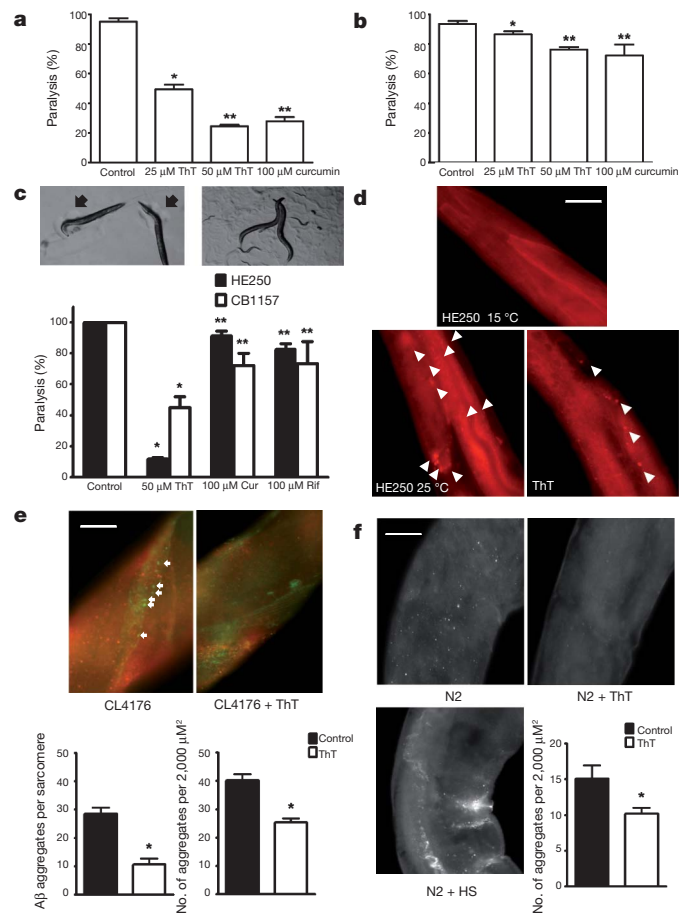

Figure 2 | ThT and curcumin rescue a paralysis phenotype and slow protein aggregation *in vivo*. a, b, The paralysis phenotype associated with protein aggregation is suppressed by 25 μM ThT, 50 μM ThT, 100 μM curcumin in CL4176 (*P < 0.001, **P < 0.0001) expressing Aβ(3–42) (**a**) and AM140 (*P < 0.05, **P < 0.01) expressing polyQ (**b**) after 1 and 8 days at 25 °C, respectively. Error bars represent the mean ± s.e.m. of four independent experiments. **c**, Temperature-sensitive strain HE250 [*unc-52(e669su250)II*] after 36 h at 25 °C showing the typical paralysis phenotype (left upper panel) and the rescue elicited by 50 μM ThT (right upper panel). Arrows indicate the halos of clearance in the bacterial lawn characteristic of paralysed worms. Lower panel shows protection (± s.e.m.) of the HE250 paralysis phenotype by 50 μM ThT, 100 μM curcumin (Cur) and 100 μM rifampicin (Rif). *P < 0.0001, **P < 0.01. n = 4 independent experiments. **d**, Perlecan immunolocalization showing disruption/aggregation pattern after 24 h at 25 °C, as compared with worms raised at the permissive temperature (upper panel), and the suppression of disruption by 50 μM ThT treatment. Sixteen of twenty worms showed similar perlecan distribution in three independent experiments. Arrows indicate perlecan aggregates. Scale bar, 30 μm. **e**, Immunolocalization of aggregation-prone soluble oligomeric protein (A11 antibody, red) and Aβ(3–42) (green) in the presence or absence of 50 μM ThT in CL4176. Scale bar, 10 μm. Error bars represent the mean ± s.e.m., 11 worms per group in 3 independent experiments. *P < 0.0001. **f**, Immunolocalization of aggregation-prone soluble oligomeric protein (A11 antibody) in the presence or absence of 50 μM ThT and under heat shock (HS) in 11 days old wild-type N2 worms. Scale bar, 20 μm. Error bars represent the mean ± s.e.m., 11 worms per group, of 3 independent experiments. *P < 0.0001. Scale bar, 10 μm.

antibody specific for such oligomers (A11) accumulated during normal ageing and after heat shock, but was significantly decreased in both CL4176 (Fig. 2e) and N2 strains (Fig. 2f) after ThT treatment, consistent with ThT affecting protein misfolding cascades.

We reasoned that ThT may also require components of the protein homeostatic network activated by DAF-16 and HSF-1 to influence protein aggregation and lifespan[8,21]. We undertook a targeted pharmacogenetic RNA interference (RNAi) screen of genes encoding several

components of the ubiquitin/proteasome system, autophagy/lysosomal machinery and molecular chaperones. First, we asked whether reducing the expression of genes encoding these proteins modulated the paralysis of metastable perlecan mutant in the absence of ThT (Fig. 3). RNAi of small chaperones known to positively modulate lifespan in *C. elegans*, HSP-16.2 and HSP-16.41 (refs 8, 22), and the mitochondrial HSP-70 (*hsp-6*) increased paralysis of the HE250 strain. An autophagy gene (*vps-34*) also influenced the HE250 mutant paralysis phenotype (Fig. 3). Interestingly, RNAi targeting *rle-1*, an E3 ubiquitin ligase that influences lifespan in *C. elegans* by determining the rate of DAF-16 degradation[23], produced a remarkable reduction in the paralysis phenotype. This led us to test *daf-16* (RNAi), but no change in the paralysis phenotype was observed, indicating that other proteins regulated by *rle-1* can influence protein homeostasis.

We then tested for interactions between these protein homeostasis factors and the protective effect elicited by ThT on the HE250 paralysis phenotype, and found that ThT protection was decreased when combined with RNAi for several stress genes (for example, *hsp-16.2*, *hsp-16.41*), consistent with a concerted action between chaperones and ThT to maintain protein conformation. Similarly, ubiquitin/proteasome (*aip-1*) and autophagy/lysosomal (*atg-9* and *vps-34*) functions were required for the beneficial effects of ThT. Interestingly, LMP-2, a protein involved in lysosome function, suppresses the ThT effect on paralysis. As *lmp-2* knockdown itself has no effect on paralysis it is possible that ThT is cleared from the cell by a lysosomal mechanism, such that *lmp-2* RNAi results in increased ThT bioactivity.

Next, we explored the dependency of DAF-16 for ThT action. The ThT suppression of the paralysis phenotype was potentiated by *daf-16* (RNAi), indicating that some proteins activated by DAF-16 interfere with the mechanism elicited by ThT to promote protein homeostasis. This is consistent with the previous report of a reduction of amyloid-β aggregation by *daf-16* (RNAi)[9]. In contrast, *skn-1* (RNAi), encoding a transcription factor that positively modulates stress resistance and longevity, was required for the ThT effect on the HE250 paralysis phenotype (Fig. 3), indicating that some SKN-1 target genes are necessary for ThT action. However, ThT does not seem to induce constitutive nuclear localization of a SKN-1 fusion protein CF2189 [Is001(*skn-1::GFP; rol-6(su1006)*)] as occurs with certain toxins (Supplementary Fig. 15).
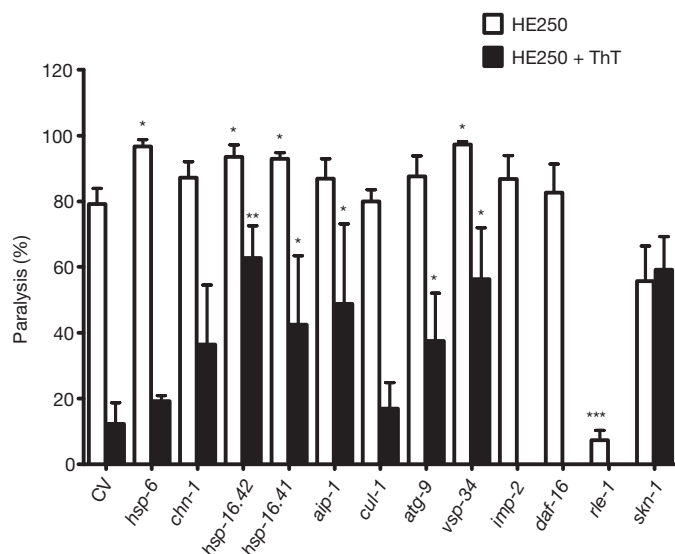

Figure 3 | Dependency of ThT suppression of protein aggregation-associated paralysis on protein homeostasis factors. RNAi by feeding was used to knockdown the expression of genes encoding proteostatic factors in HE250 in the presence or absence of 50 μM ThT and the paralysis phenotype was scored after 36 h. Proportion of worms paralysed is plotted (mean ± s.e.m.). *$P < 0.01$, **$P < 0.001$, ***$P < 0.0001$ versus control vector (CV).
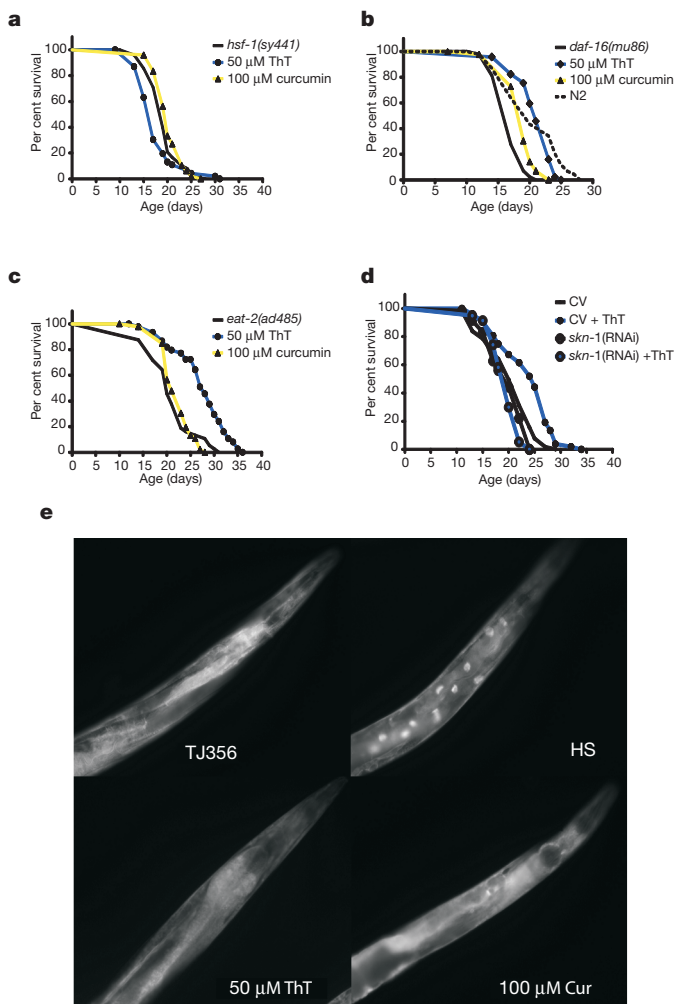

Figure 4 | ThT enhancement of lifespan depends on HSF-1 and SKN-1 transcription factors but not on DAF-16. **a–c**, Effect of 50 μM ThT and 100 μM curcumin on Kaplan–Meier survival curves of synchronously ageing PS3551 [*hsf-1(sy441)I*] (**a**), CF1038 [*daf-16(mu86)I*] (**b**) and DA465 [*eat-2(ad465)II*] (**c**) lifespan. **d**, Effect of reducing SKN-1 by *skn-1* RNAi on the increase in lifespan elicited by ThT. Plots are representative of three independent experiments. **e**, ThT and curcumin (Cur) treatment does not result in DAF-16::GFP relocalization as compared to control strain [*daf-16::daf-16-gfp + rol-6*] (upper left). Control strain under heat shock (HS) produces a clear relocalization of DAF-16::GFP to the nuclei of intestinal cells.

We returned then to the question of whether ThT was extending lifespan by a mechanism related to protein homeostasis and initially focused on the transcription factor genes *hsf-1* and *daf-16*; mutation of either one shortens normal lifespan and suppresses the beneficial effects of a *daf-2* mutation on the amyloid-β-aggregation model in *C. elegans*[9]. We found that ThT does not increase lifespan in a strain carrying the *hsf-1(sy441)* mutation—resulting in non-functional HSF protein (Fig. 4a and Supplementary Fig. 11)—indicating that the ThT effect on lifespan requires the participation of HSF-1-regulated machinery. Consistent with this idea, we found that the protein levels of HSP-16.2 and HSP-70, the mRNA levels of a mitochondrial (*hsp-6*) and a cytosolic (*chn-1*) *hsp-70* isoform are upregulated by ThT treatment (Supplementary Fig. 12a). We also detected a slight increase in the levels of HSF-1 protein under ThT treatment (Supplementary Fig. 12b).

ThT treatment did extend the lifespan of *daf-16(mu86)* worms lacking functional DAF-16 (Fig. 4b) and in a long-lived ILS mutant, *age-1(hx546)*, which is hypomorphic for the p110 catalytic subunit of a phosphoinositide 3-kinase (Supplementary Fig. 13). In addition, ThT treatment did not alter the localization of a DAF-16 fusion protein TJ356 (zIs356 IV [*daf-16::daf-16-gfp; pRF4 (rol-6(su1006))*]), in contrast

to the effect of stress (Fig. 4e; Supplementary Fig. 14). We conclude that ThT lifespan extension is independent of ILS.

Because ThT suppression of the paralysis phenotype on the HE250 strain required the *skn-1* transcription factor, we tested if it was required for lifespan extension. We found that ThT does not increase the lifespan of nematodes exposed to *skn-1* RNAi and conclude that this transcription factor is, along with HSF-1, required for the increase in lifespan elicited by ThT (Fig. 4d).

We have observed that compounds traditionally used to stain amyloid-β deposits confer a large increase in lifespan to *C. elegans*. ThT is capable of suppressing protein-aggregation-associated paralysis in toxic protein models in multiple tissues. ThT reduces Aβ(3–42) aggregation, decreases the levels of soluble aggregation-prone oligomeric proteins and localizes with these aggregates *in vivo*. The mechanism of aggregation suppression depends on molecular chaperones, autophagy and proteosomal functions. Finally, the extent of the ThT-mediated lifespan increase depends on the transcription factors HSF-1 and SKN-1. We propose that amyloid-binding compounds act as stress response mimetics and activate the *C. elegans* stress-response pathways regulated by the HSF-1 and SKN-1 transcription factors, leading to stabilization of misfolded proteins and increased lifespan. Given the known ability of these compounds to bind to amyloid, we propose that they may also directly interact with aggregating proteins, promoting proper folding and possibly activating stress response pathways by an unknown mechanism (Supplementary Fig. 1). This modulation of protein homeostasis and protein aggregation pathways has beneficial effects for healthspan and lifespan. Small stress response mimetic molecules that target protein homeostatic mechanisms may provide opportunities for intervention in ageing and age-related disease.

METHODS SUMMARY

Lifespan assays were performed on nematode growth media plates with the compounds spread over the surface of the plates and 60 µl of a concentrated suspension of *Escherichia coli* OP50 in the presence of FUdR (75 µM) at 20 °C. Nematodes were scored as alive, dead or lost every other day. Nematodes that failed to display touch-provoked movement were scored as dead. For dietary restriction experiments, bacterial concentration was adjusted to 1.0×10^9 c.f.u. ml^{-1}, spotted on peptone-free 2% agar in the presence or absence of the specified compounds. Estimates of initial mortality rate, rate of increase with age and model fitting to Gompertz mortality curves were made using WinModest. Scoring for paralysis was initiated 1, 2 and 8 days after temperature upshift for HE250, CL4176 and AM140, respectively. Nematodes were scored as paralysed if they exhibited 'halos' of cleared bacteria around their heads, eggs accumulated close to the body or if they failed to respond to a touch-provoked movement. For fluorescence microscopy, strains were paralysed with 1 mM levamisole, mounted on 1% agarose pads and imaged. For immunofluorescence, worms were collected after treatment, rinsed, fixed in 4% paraformaldehyde overnight, permeabilized by 24 h exposure to β-mercaptoethanol at 37 °C followed by collagenase treatment and processed for different antigens. RNAi bacterial strains expressing double-stranded RNA that inactivates specified genes were used to knockdown the expression of specific genes of interest. For real-time quantitative PCR with reverse transcription (RT–qPCR) analysis, RNA from individual worms was extracted and the sample was added to the reverse transcription reaction and reversed transcribed at 50 °C for 20 min. mRNAs were found to have invariant steady-state levels across treatments and were used to derive calibrated normalized relative quantities (CNRQs) for each gene of interest. Differences in relative mRNA transcript levels were identified using pair-wise *t*-tests.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Morimoto, R. I. Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. *Genes Dev.* **22**, 1427–1438 (2008).
2. Balch, W. E., Morimoto, R. I., Dillin, A. & Kelly, J. W. Adapting proteostasis for disease intervention. *Science* **319**, 916–919 (2008).
3. Porat, Y., Abramowitz, A. & Gazit, E. Inhibition of amyloid fibril formation by polyphenols: structural similarity and aromatic interactions as a common inhibition mechanism. *Chem. Biol. Drug Des.* **67**, 27–37 (2006).
4. Frid, P., Anisimov, S. V. & Popovic, N. Congo red and protein aggregation in neurodegenerative diseases. *Brain Res. Brain Res. Rev.* **53**, 135–160 (2007).
5. Kenyon, C. The plasticity of aging: insights from long-lived mutants. *Cell* **120**, 449–460 (2005).
6. Tullet, J. M. *et al.* Direct inhibition of the longevity-promoting factor SKN-1 by insulin-like signaling in *C. elegans*. *Cell* **132**, 1025–1038 (2008).
7. Morley, J. F. & Morimoto, R. I. Regulation of longevity in *Caenorhabditis elegans* by heat shock factor and molecular chaperones. *Mol. Biol. Cell* **15**, 657–664 (2004).
8. Hsu, A. L., Murphy, C. T. & Kenyon, C. Regulation of aging and age-related disease by DAF-16 and heat-shock factor. *Science* **300**, 1142–1145 (2003).
9. Cohen, E., Bieschke, J., Perciavalle, R. M., Kelly, J. W. & Dillin, A. Opposing activities protect against age-onset proteotoxicity. *Science* **313**, 1604–1610 (2006).
10. Groenning, M. Binding mode of Thioflavin T and other molecular probes in the context of amyloid fibrils—current status. *J. Chem. Biol.* **3**, 1–18 (2009).
11. Rodríguez-Rodríguez, C. *et al.* Design, selection, and characterization of thioflavin-based intercalation compounds with metal chelating properties for application in Alzheimer's disease. *J. Am. Chem. Soc.* **131**, 1436–1451 (2009).
12. Drake, J., Link, C. D. & Butterfield, D. A. Oxidative stress precedes fibrillar deposition of Alzheimer's disease amyloid β-peptide (1–42) in a transgenic *Caenorhabditis elegans* model. *Neurobiol. Aging* **24**, 415–420 (2003).
13. McColl, G. *et al.* The *Caenorhabditis elegans* Aβ$_{1-42}$ model of Alzheimer's disease predominantly expresses Aβ$_{3-42}$. *J. Biol. Chem.* **284**, 22697–22702 (2009).
14. Temussi, P. A., Masino, L. & Pastore, A. From Alzheimer to Huntington: why is a structural understanding so difficult? *EMBO J.* **22**, 355–361 (2003).
15. Gidalevitz, T., Ben-Zvi, A., Ho, K. H., Brignull, H. R. & Morimoto, R. I. Progressive disruption of cellular protein folding in models of polyglutamine diseases. *Science* **311**, 1471–1474 (2006).
16. Zengel, J. M. & Epstein, H. F. Identification of genetic elements associated with muscle structure in the nematode *Caenorhabditis elegans*. *Cell Motil.* **1**, 73–97 (1980).
17. Anderson, P. & Brenner, S. A selection for myosin heavy chain mutants in the nematode *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **81**, 4470–4474 (1984).
18. Ben-Zvi, A., Miller, E. A. & Morimoto, R. I. Collapse of proteostasis represents an early molecular event in *Caenorhabditis elegans* aging. *Proc. Natl Acad. Sci. USA* **106**, 14914–14919 (2009).
19. Lakowski, B. & Hekimi, S. The genetics of caloric restriction in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **95**, 13091–13096 (1998).
20. Chen, D., Thomas, E. L. & Kapahi, P. HIF-1 modulates dietary restriction-mediated lifespan extension via IRE-1 in *Caenorhabditis elegans*. *PLoS Genet.* **5**, e1000486 (2009).
21. Murphy, C. T. *et al.* Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. *Nature* **424**, 277–283 (2002).
22. Walker, G. A. & Lithgow, G. J. Lifespan extension in *C. elegans* by a molecular chaperone dependent upon insulin-like signals. *Aging Cell* **2**, 131–139 (2003).
23. Li, W., Gao, B., Lee, S. M., Bennett, K. & Fang, D. RLE-1, an E3 ubiquitin ligase, regulates *C. elegans* aging by catalyzing DAF-16 polyubiquitination. *Dev. Cell* **12**, 235–246 (2007).

METHODS

Nematode growth and strains. Strains were cultured under standard laboratory conditions in USP agar. Strains used in this work include N2, HE250 [*unc-52(e669su250)II*], CB1157 [*unc-54(e1157)I*], CF1038 [*daf-16(mu86)I*], DA465 [*eat-2(ad465)II*], TJ1052 [*age-1(hx546)II*], PS3551 [*hsf-1(sy441)I*], TJ356 [*zIs356 IV (daf-16::daf-16-gfp; pRF4 (rol-6(su1006))*], CL4176 [*dvIs27[myo-3::Aβ(3–42)-let 3′UTR(pAF29); pRF4 (rol-6(su1006))]*], AM140 (*rmIs132[P(unc-54) Q35::YFP]*), ZZ26 (*unc-63(x26)I*), CW152 (*gas-1(fc21) X*), CF2189 (*Is001[skn-1::gfp rol-6(su1006)]*).

Lifespan assay. Lifespan assays were performed as described previously[24]. Briefly, the nematode growth media (NGM, USP agar) plates were prepared under sterile conditions. One-hundred microlitres of concentrated stocks of each of the compounds used in this study were added onto a previously prepared NGM small plate (3 ml volume) immediately spread over the surface of the plate. The final concentrations quoted in the text assume an even distribution of compound throughout the 3 ml of agar in the plate. The plates were then placed in a laminar flow hood at room temperature (22 °C) for 30 min and then 60 μl of a concentrated suspension of *E. coli* OP50 was spotted to form a circular lawn on the centre of each plate. Thirty late L4 larvae growing at 20 °C were transferred to fresh NGM plates with FUdR (75 μM, unless otherwise stated) in the presence or absence of the specified compounds and incubated at 20 °C. The first day of adulthood is day 3 in survival curves. We noted between-experiment variation in the magnitude of the lifespan extension observed with ThT, which appeared to correlate with different suppliers and batches. ThT concentration should be optimized depending on batch and purity and stability of the compound. We noted that a darkening in the appearance of the stock resulted in loss of lifespan extension activity and even early death. The optimal range for lifespan extension was between 25 and 75 μM.

Nematodes were scored as alive, dead or lost every second day. Nematodes that failed to display touch-provoked movement were scored as dead. Nematodes that died from causes other than ageing, such as sticking to the plate walls, internal hatching of eggs ('bagging') or gonadal extrusion were censored as lost worms. Nematodes were transferred to fresh plates every 3–6 days. All lifespan experiments were performed at 20 °C unless otherwise stated. Survival curves were plotted and statistical analyses (log-rank test) were performed using Prism 4 software (Graphpad Software).

Dietary restriction. Plates were prepared as described[25] but bacterial concentration was adjusted to 1.0×10^{12} c.f.u. ml^{-1} and diluted to achieve a bacterial concentration of 1.0×10^9 c.f.u. ml^{-1}. Diluted bacterial cultures were spotted onto dietary restriction agar plates, which were modified from the standard nematode growth media (NGM) plates by excluding peptone and increasing agar from 1.7% to 2.0%. Carbenicillin (50 mg ml^{-1}) was added to the agar plates to further prevent bacterial growth. Synchronized L4 larvae grown under standard laboratory conditions (NGM plates with OP50 food, 20 °C) were transferred to fresh dietary restriction agar plates in the presence or absence of 1, 10, 25, 50 and 100 μM ThT, and lifespans scored as described earlier.

Demographic analysis. Estimates of initial mortality rate and rate of increase with age and model fitting were made using WinModest. Gompertz mortality curves, ln(ux) = ln(a) + bx, where ux defines the age-specific hazard, were fitted with log-likelihood ratios used to examined the effects of constraining the intercept (a) or gradient (b) variables.

Worm paralysis assays. Populations of CL4176 *dvIs2[pCL12(unc-54/human Aβ₃–₄₂ minigene) + pRF4]* or AM140 (*rmIs132[P(unc-54) Q35:YFP]*) worms were grown at 20 °C for 48 h and then exposed to 50 μM ThT and 100 μM curcumin at 25 °C in presence of FUdR (10 μg ml^{-1}) for AM140. Scoring for paralysis was initiated 2 and 8 days after temperature upshift for CL4176 and AM140, respectively. Nematodes were scored as paralysed if they failed to move during observation and exhibited 'halos' of cleared bacteria around their heads (indicative of insufficient body movement to access food), eggs accumulated close to the body or if they failed to respond to a touch-provoked movement with a platinum wire. For sensitivity to ethanol or levamisole resistance, CW152 *gas-1(fc21 X)* and ZZ26 (*unc-63(x26)I*) worms were picked into 0.4 M ethanol or 50 μM levamisole, respectively, equilibrated for 5 min, and scored for paralysis as described earlier. Treated and untreated worms were compared with an unpaired *t*-test (implemented in Prism 4, Graphpad Software).

Immunostaining and photomicroscopy. TJ356 [*zIs356 IV (daf-16::daf-16-gfp; pRF4 (rol-6(su1006))*] or CF2189 *Is001(skn-1::gfp + rol-6(su1006))* nematodes were paralysed with 1 mM levamisole mounted on 1% agarose pads and imaged using Olympus BX51 (×60 objective) and HCImage software (Hamamatsu). For positive controls, TJ356 worms were exposed for 3 h to 33 °C and CF2189 worms were exposed to 5 mM NaAsO₂ for 15 min. For immunofluorescence, N2, HE250 *unc-52(e669su250)*, CB1157 *unc-54(e1157)* or CL4176 *dvIs27[myo-3::Aβ(3–42)-let 3′UTR(pAF29); pRF4 (rol-6(su1006))]* worms were treated for 24–36 h with or without 50 μM ThT at 25 °C. After this period the worms were collected, rinsed and fixed in 4% paraformaldehyde overnight. After fixation, worms were rinsed twice with 1 ml of 10 mM Tris-HCl pH 7.5 and then permeabilized by 24 h exposure to β-mercaptoethanol at 37 °C followed by collagenase treatment (2 mg ml^{-1} for 1–1.5 h at 37 °C) to allow for digestion of the cuticle. Paramyosin and perlecans were detected with primary monoclonal antibodies 5-23 and MH3

(developed by H. Epstein and R. H. Waterston and obtained from the Developmental Studies Hybridoma Bank developed under the auspices of the NICHD and maintained by The University of Iowa, Department of Biology) and AlexaFluor 633 goat anti-mouse (Molecular Probes) as secondary antibody. Soluble oligomers and amyloid-β peptide were detected with anti polyclonal A11 (Invitrogen) and 6E10 monoclonal (Covance) primary antibiodies, respectively, with AlexaFluor 568 goat anti-rabbit and AlexaFluor 488 goat anti-mouse (Molecular Probes) as secondary antibodies. Image analysis was performed in Image Analyst MKII 2.0.49 (Image Analyst Software) as follows. Aggregates were counted by image segmentation. To this end, wide-field epifluorescence and confocal micrographs were pre-processed by high pass Butterworth filtering at $\omega_{cuton} = 0.85$ cycles μm^{-1} (order 1.5) to amplify small (<1.2 μm) punctate details, smoothed by Wiener filtering and rescaled with a gamma level 0.7. Image segmentation was performed by a modified Watershed method from seeds defined by being brighter than the 99.9 percentile of the rescaled image. Pixels brighter than 10% of the peak intensity for each aggregate were taken as positive. Because the high-pass filtering subtracts local mean, the segmentation resulted in objects outlined close to their original half maximal intensities, and therefore the size of segmented objects truly reflected the size of the aggregates. Objects larger than 30 pixels (0.56 μm^2) were rejected. The high-pass filtering also eliminated the blur of wide-field images, therefore these images were handled in the same way as confocal images. Images of different conditions were handled with the same algorithm, and no subjective threshold levels were applied. The number of aggregates was determined in same-size areas close to the vulva of at least 10 worms under the same conditions.

We explored ThT distribution and potential colocalization with proteins prone to aggregate by using two-photon excitation of ThT at 800 nm and emission at 435–485 nm in combination with anti-oligomers and amyloid-β peptide immunodetection described earlier. In this spectral range worms exhibited negligible autofluorescence, therefore the signal was highly specific for ThT. Considering that image acquisition was performed after immunostaining probably only the protein-bound form of ThT was imaged. Pearson's coefficient values for ThT and amyloid-β were calculated by using Image Analyst MKII 2.0.49 (Image Analyst Software).

Western blot. Peptide corresponding to amino acids 110–145 (NLSEDGKL SIEAPKKEAVQGRSIPIQQAIVEEKSAE) of HSP-16.2 was used to commercially synthesize antiserum (Invitrogen). Briefly, KLH-peptide was emulsified by mixing with an equal volume of Freund's adjuvant and injected into three subcutaneous dorsal sites for a total of 0.1 mg of peptide for immunization. The animals (rabbits) were bled, the blood allowed to clot and the serum collected by centrifugation. Monoclonal HSP-70 and polyclonal HSF-1 primary antibodies were from Stressgen (N27F3-4 and SPA-901, respectively).

For immunoblot analysis, 3-day-old adult hermaphrodites were treated with 50 μM ThT or 100 μM curcumin as described earlier and replicates of 25 nematodes were collected for each treatment. Worms were transferred to siliconized eppendorf tubes, washed once in S-basal and frozen in liquid N₂. Standard SDS–PAGE was performed using (4–12%) NOVEX gels and MES running buffer. Following transfer PVDF (BioRad) membranes were incubated with antisera (1:10,000) or primary antibodies (1:1,000) diluted in blocking buffer and then with secondary, goat anti-rabbit IgG antibody/horseradish peroxidase conjugate (Pierce), diluted 1:25,000. Detection was undertaken with chemiluminescent reagents (SuperSignal, Pierce) and standard autoradiography.

RNAi knockdown of gene expression. RNAi bacterial strains expressing double-stranded RNA that inactivates specified genes were cultured and used as previously described[26]. Briefly, eggs isolated from synchronous populations of cultures were placed on fresh RNAi plates and allowed to grow at 15 °C; 3 days later, L4 moult nematodes were transferred to new plates seeded with the same bacteria in the presence or absence of compounds and switched to 25 °C. In all cases, 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) was used for induction of double-stranded RNA. In all cases the identity of the clones was confirmed by sequencing.

Real-time qPCR analysis. Twelve single adults from control or 50 μM ThT populations were picked after 3, 6 and 12 days of treatment into 5 μl of distilled water and flash frozen until extraction. Individual worms were extracted and analysed on the QIAcube robot as described previously[27]. The housekeeping genes *gpd-1* and *gpd-4* were found to have invariant steady-state levels across treatments and were used to derive calibrated normalized relative quantities (CNRQs) for each gene of interest as previously described[27].

24. McColl, G. *et al.* Pharmacogenetic analysis of lithium-induced delayed aging in *Caenorhabditis elegans*. *J. Biol. Chem.* **283**, 350–357 (2008).
25. Chen, D., Thomas, E. L. & Kapahi, P. HIF-1 modulates dietary restriction-mediated lifespan extension via IRE-1 in *Caenorhabditis elegans*. *PLoS Genet.* **5**, e1000486 (2009).
26. Timmons, L., Court, D. L. & Fire, A. Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* **263**, 103–112 (2001).
27. McColl, G. *et al.* Insulin-like signaling determines survival during stress via posttranscriptional mechanisms in *C. elegans*. *Cell Metab.* **12**, 260–272 (2010).

LETTER

Structure of mammalian AMPK and its regulation by ADP

Bing Xiao[1]*, Matthew J. Sanders[2]†*, Elizabeth Underwood[1]*, Richard Heath[1]*, Faith V. Mayer[2]*, David Carmena[2], Chun Jing[1], Philip A. Walker[1], John F. Eccleston[1]‡, Lesley F. Haire[1], Peter Saiu[1]†, Steven A. Howell[1], Rein Aasland[1]†, Stephen R. Martin[1], David Carling[2] & Steven J. Gamblin[1]

The heterotrimeric AMP-activated protein kinase (AMPK) has a key role in regulating cellular energy metabolism; in response to a fall in intracellular ATP levels it activates energy-producing pathways and inhibits energy-consuming processes[1]. AMPK has been implicated in a number of diseases related to energy metabolism including type 2 diabetes, obesity and, most recently, cancer[2–6]. AMPK is converted from an inactive form to a catalytically competent form by phosphorylation of the activation loop within the kinase domain[7]: AMP binding to the γ-regulatory domain promotes phosphorylation by the upstream kinase[8], protects the enzyme against dephosphorylation, as well as causing allosteric activation[9]. Here we show that ADP binding to just one of the two exchangeable AXP (AMP/ADP/ATP) binding sites on the regulatory domain protects the enzyme from dephosphorylation, although it does not lead to allosteric activation. Our studies show that active mammalian AMPK displays significantly tighter binding to ADP than to Mg-ATP, explaining how the enzyme is regulated under physiological conditions where the concentration of Mg-ATP is higher than that of ADP and much higher than that of AMP. We have determined the crystal structure of an active AMPK complex. The structure shows how the activation loop of the kinase domain is stabilized by the regulatory domain and how the kinase linker region interacts with the regulatory nucleotide-binding site that mediates protection against dephosphorylation. From our biochemical and structural data we develop a model for how the energy status of a cell regulates AMPK activity.

At the whole-body level, AMPK is regulated by a diverse range of hormones—for example, leptin[10], adiponectin[11], ciliary neurotrophic factor[12] and ghrelin[13]—and it has a role in appetite[13,14], glucose, lipid and protein metabolism[1,3,15], cell growth, and cell polarity[2,4]. AMPK is a heterotrimeric complex comprising an α-catalytic subunit and two regulatory subunits (β and γ)[1]. Activation of AMPK requires phosphorylation of Thr 172, which lies in the activation segment of the amino-terminal kinase domain of the α-subunit[7]. Phosphorylation of Thr 172 leads to a several-hundred-fold increase in activity[9,16]. In mammals, calcium/calmodulin-dependent protein kinase kinase-β (CaMKKβ) and LKB1 are the predominant kinases upstream of AMPK, although there is some evidence implicating other upstream kinases[17,18]. Previous studies have shown that AMP protects against the dephosphorylation of Thr 172 (refs 16, 19) and we recently provided evidence that protection against dephosphorylation is the major physiological mechanism for activation of AMPK[9]. In addition to activation by phosphorylation, AMP causes a 2–5-fold allosteric activation of AMPK depending on the nature of the isoforms present in the AMPK complex[20]. To address this issue we previously investigated the nucleotide-binding properties of the γ1 subunit of AMPK and determined the structure of the regulatory core of mammalian AMPK (C-terminal domain of α-subunit, C-terminal domain of β-subunit, full-length γ-subunit; hereafter referred to as the regulatory fragment) in complex with nucleotides[21]. Notably, our studies revealed that three of the four potential nucleotide-binding sites are occupied[21]. One of these sites contains a permanently bound AMP molecule (site 4, following nomenclature proposed by ref. 22), whereas AMP and Mg-ATP compete for binding at the other two sites[21] (site 1 and site 3).

Unlike AMP, ADP has no significant allosteric effect on AMPK isolated from rat liver[23]. Consistent with this, we also found that ADP does not activate recombinant AMPK under conditions where AMP produces a twofold activation (Fig. 1a). However, our studies show that ADP provides protection of AMPK from dephosphorylation across a similar range of concentrations as AMP (Fig. 1b). We have also shown the same effect using AMPK purified from rat liver (Supplementary Fig. 2a). Although Mg-ATP does not protect against dephosphorylation (Fig. 1c), it does compete with the protective effect of both AMP and ADP on dephosphorylation (Fig. 1d). We have also shown that the protective effect of ADP is lost in a Wolff–Parkinson–White syndrome mutation (Supplementary Fig. 2b). We propose that AMP or ADP (AMP/ADP) binding shifts the equilibrium between dephosphorylation-sensitive and -insensitive states, and thus slows, but does not abolish, dephosphorylation of the enzyme by phosphatases.

Extending our earlier work looking at nucleotide binding to the regulatory fragment, we characterized binding of nucleotides to active full-length AMPK. For these studies we used CaMKKβ to stoichiometrically phosphorylate Thr 172 on the activation loop of recombinant full-length AMPK, we used the coumarin adducts of ATP and ADP as fluorescent reporters of nucleotide binding, and derived the binding parameters for the unlabelled nucleotides by competition experiments (Fig. 2a). We verified that the two species bind at the same sites by determining the crystal structures of the regulatory fragment in complex with coumarin-ADP and with ADP (Supplementary Fig. 3). The results show (Table 1) that the two exchangeable sites have markedly different affinities for nucleotides. Binding at the tighter of the two sites is at least 30-fold stronger than at the weaker site. Given that under physiological conditions most of the ATP is coordinated to Mg^{2+}, and that the majority of AMP and ADP is not, we also measured nucleotide binding in the presence of this cation. The data show that Mg-ATP binds up to tenfold weaker than ATP. Thus, active AMPK binds AMP/ADP significantly more strongly than it does Mg-ATP at both exchangeable sites. There are two lines of evidence that lead us to conclude that it is AMP/ADP binding at the weaker of the two exchangeable sites that accounts for the protection of the enzyme against dephosphorylation. The first is that the dose-response curve for AMP/ADP-mediated protection against phosphorylation correlates with the binding curves for these nucleotides at the weaker, rather than the stronger, of the two binding sites (Fig. 1b). The second comes from our discovery that NADH binds to AMPK.

[1]MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK. [2]MRC Clinical Sciences Centre, Hammersmith Hospital Campus, Imperial College, DuCane Road, London W12 0NN, UK. †Present addresses: MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK (M.J.S.); Department of Molecular Biology, University of Bergen, Thormöhlensgt. 55, N5020 Bergen, Norway (R.A.); Cancer Research UK, 44 Lincoln's Inn Fields, PO Box 123, London WC2A 3PX, UK (P.S.).
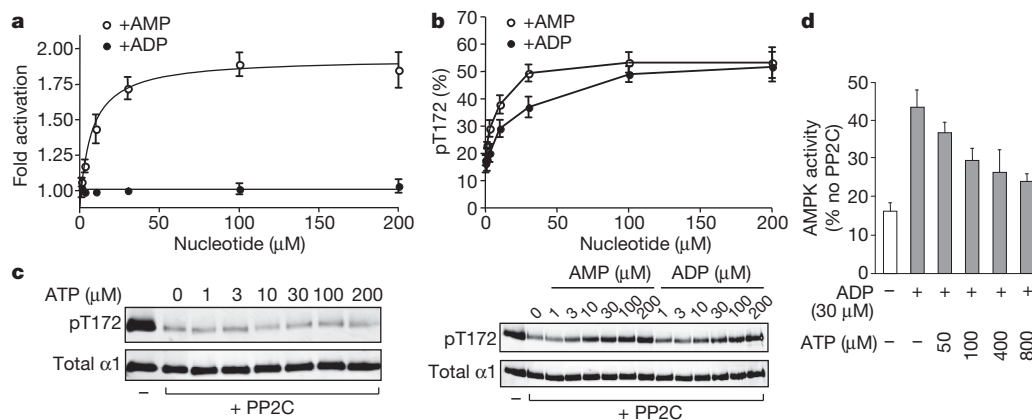*These authors contributed equally to this work.
‡Deceased.

Figure 1 | Role of ADP in regulation of AMPK activity.
a, AMP, but not ADP, allosterically activates AMPK. **b**, AMP and ADP protect AMPK from dephosphorylation. **c**, ATP does not protect against dephosphorylation. **d**, Mg-ATP competes with the protective effect of ADP on dephosphorylation. All results are displayed as the mean ± s.e.m. determined from at least three independent experiments. Where appropriate a representative blot (*n* = 3) showing Thr 172 phosphorylation and total α-subunit levels is shown.

NADH undergoes a significant change in fluorescence upon binding to AMPK. We used this property to establish that the cofactor binds to a single site on the enzyme, with a dissociation constant (K_d) of about 50 μM (Fig. 2b, inset). NADH binding is competed by AXPs binding to the stronger, but not the weaker, of the two exchangeable sites (Fig. 2b and Table 1). When we repeated the ADP protection against dephosphorylation experiments using a range of NADH concentrations, we found no evidence for NADH competing with the protective effect of ADP on dephosphorylation, whereas NADH and ADP both compete with AMP for allosteric activation of the enzyme (Supplementary Fig. 4). This observation indicates that it is AMP/ADP binding at the weaker of the two exchangeable sites, the one that does not bind NADH, that is responsible for protection against dephosphorylation. We also carried out co-crystallization of the regulatory fragment with one molar equivalent of ADP (Supplementary Fig. 5). The resulting electron density map showed full occupancy of ADP at site 1 and no detectable density at site 3, identifying site 3 as the weaker binding site. We can therefore assign the allosteric effect to AMP binding at the tighter site 1, and protection against dephosphorylation is mediated by AMP/ADP binding at the weaker site 3.

Previous studies on the regulation of AMPK have focused on the role of AMP because it allosterically activates the enzyme[15] whereas ADP does not. However, phosphorylation remains central to AMPK regulation as the enzyme is inactive in the absence of Thr 172 phosphorylation[7,16]. Under optimal conditions, mammalian cells maintain ATP at a high level relative to ADP and AMP. Typical concentrations

of free adenine nucleotides in mammalian cells lie in the range of 3,000–8,000 μM for ATP, 50–200 μM for ADP and 0.5–5 μM for AMP[24–26]. Because the free concentration of ADP is between 10- to 400-fold higher than AMP, and their binding constants are similar, ADP will be more successful at competing with Mg-ATP than AMP. Therefore, the fact that ADP protects AMPK from dephosphorylation is likely to represent an important physiological mechanism for regulating the activity of the enzyme.

We have also determined the crystal structure of an active form of the enzyme that encompasses the whole of the catalytic α-subunit. The construct used is shown in Fig. 3a (details of its design are given in Supplementary Fig. 6). From the best samples we collected a data set to 3.3 Å Bragg spacing, after screening about 100 crystals, and solved the structure by molecular replacement using independent models for the regulatory fragment (Protein Data Bank 2V8Q)[21] and the kinase domain (Protein Data Bank 2H6D)[27]. Although the data set is at medium resolution, the molecular replacement solution was robust and yielded initial electron density that revealed the location of many components that were not present in the original model. As might be expected for a structure of this complexity and resolution, some parts of the molecule are better defined than others. For example, the activation loop of the kinase domain, which is packed against the regulatory fragment, has better defined electron density than loops on the surface of the complex, which often show continuous main-chain density but lack side-chain features. Overall, the most important features of the current structure concern the architecture of the complex, particularly the relationships between the α-kinase domain and the α-linker with the regulatory fragment (omit maps for these regions are presented in Supplementary Fig. 7).

The structure is shown in Ribbons representation (Fig. 3b, c and Supplementary Fig. 8) and in space-filling representation before and after the kinase domain and the linker region have been rotated away from the complex to display their contact regions (Fig. 3d). The first of these interfaces involves the kinase domain and has a surface area of about 1,100 Å^2. This relatively modest contact area is consistent with the observation that specific protease cleavage of the linker between the kinase domain and the C-terminal domain of the α-subunit leads to material that separates into two components (kinase domain plus regulatory fragment) after gel filtration. A significant part of the
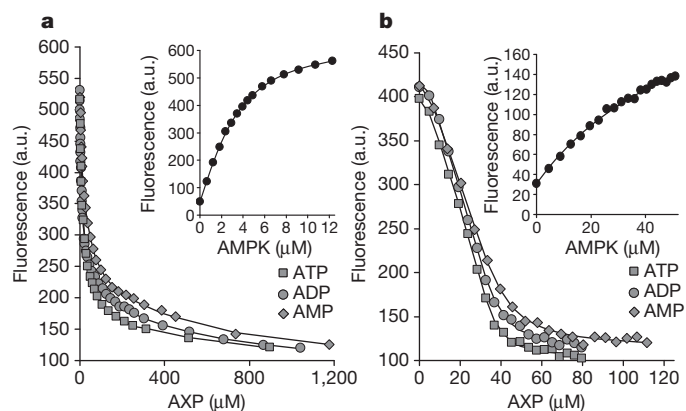

Figure 2 | Measurement of equilibrium dissociation constants for the binding of AXPs to phosphorylated AMPK. **a**, Displacement of coumarin-ATP (C-ATP) from the AMPK–(coumarin-ATP)$_2$ complex by AXPs monitored using fluorescence at 470 nm. Solid lines are the computed best fits with $K_{d,I}$ and $K_{d,II}$ for C-ATP binding to AMPK fixed at 1.1 and 4.2 μM. Inset: titration of coumarin-ATP with AMPK. **b**, Displacement of NADH from the AMPK–NADH complex by AXPs monitored using fluorescence at 435 nm. The solid line is the computed best fit with the K_d for NADH fixed at 65 μM. Inset: fluorescence titration of NADH with AMPK.

Table 1 | Equilibrium K_d values for the binding of AXPs to phosphorylated AMPK

Ligand	K_d (μM) versus NADH	$K_{d,I}$ (μM)	$K_{d,II}$ (μM)
		versus C-AXPs	
AMP	1.6 (0.5)	2.5 (0.6)	80 (25)
ADP	1.3 (0.5)	1.5 (0.4)	50 (15)
ATP	0.9 (0.3)	1.7 (0.5)	65 (15)
Mg-ATP	32 (12)	18 (7.5)	230 (80)

Dissociation constants (K_d) were determined at 20 °C by competition against NADH or coumarin-AXP (C-AXPs) in 25 mM Tris, 1 mM TCEP, 100 mM NaCl (pH 8) with and without 5 mM MgCl$_2$. The K_d values are reported as the mean (±s.d.) determined from at least three independent measurements.

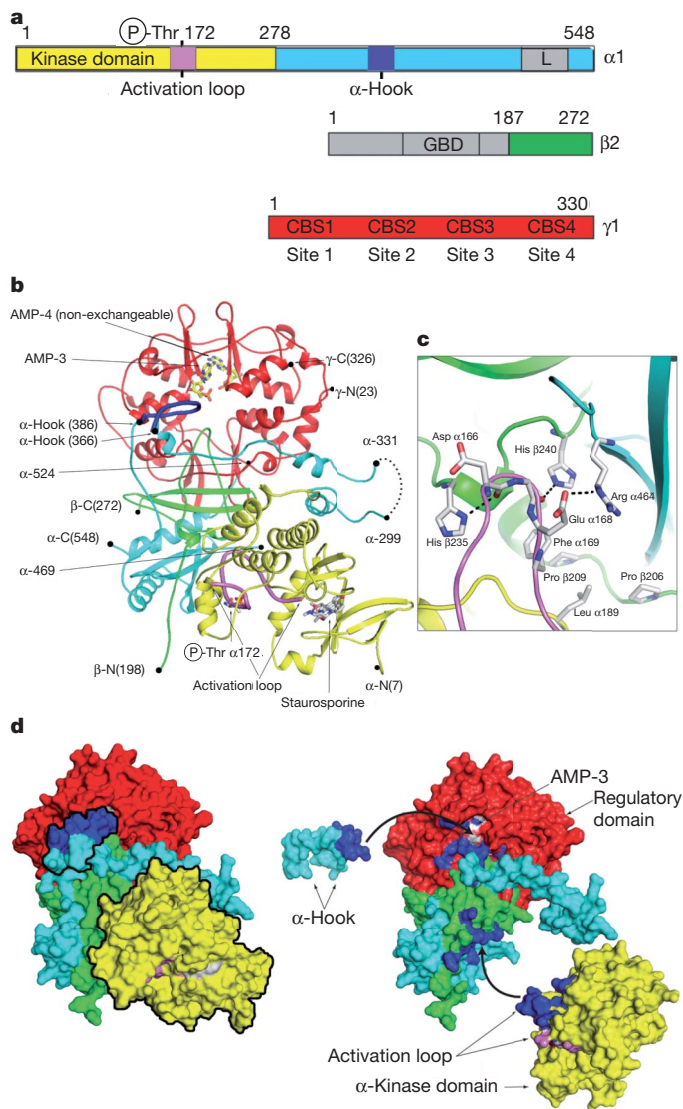**Figure 3 | Crystal structure of active mammalian AMPK. a,** Schematic representation of the components of the heterotrimer; the parts of the complex missing from the crystallized protein are shown in grey. The domains, including the activation loop (pink) and α-hook (dark blue), are coloured the same in all panels. GBD, glycogen-binding domain. **b,** Ribbon representation of the crystallized complex with two bound AMPs, staurosporine and phosphorylated Thr 172 shown in stick representation. The α-hook and activation loop of the kinase domain are shown in heavier lines and coloured dark blue and pink, respectively. **c,** The interface between the activation loop and the regulatory fragment is shown in more detail in a similar orientation as **b**; potential electrostatic interactions are indicated by dashed lines. **d,** The complex is shown in two space-filling representations. The left panel represents the same view as **b** with the α-hook and kinase domain outlined in black. In the right-hand panel these two components have been rotated away from the regulatory domain to show the interaction surfaces in an 'open-book' representation, where the contacting residues have been coloured in dark blue. With the α-hook removed, AMP-3 becomes visible. The black arrows indicate the rotations that would reassemble the complex.

contact area involves the activation loop of the kinase interacting predominantly with the C terminus of the β-domain (Fig. 3c). Unlike previous structures of the isolated and non-phosphorylated kinase domain, the phosphorylated activation loop in our structure is well ordered, as evidenced by electron density maps (Supplementary Fig. 7). The small lobe of the kinase is in a more closed conformation relative to the unphosphorylated (and thus inactive) isolated kinase domain structure[27] (Supplementary Fig. 9). The fact that the activation loop mediates the interaction of the kinase domain with the regulatory

fragment means that, in this conformation, Thr 172 is protected from access by phosphatases. This idea is strongly supported by site-directed mutagenesis experiments: mutation of β1 His 233 (corresponding to His-235 in β2) at the interface with the kinase domain (Fig. 3c) results in an enzyme that is activatable by phosphorylation but that has a significantly increased rate of dephosphorylation in phosphatase assays (Fig. 4a, b).

Another component of the α-subunit–regulatory fragment interaction is provided by a part of the segment of the α-chain that links the N-terminal kinase domain to the C-terminal regulatory fragment, involving residues between α373 and α382 that are largely conserved between α1 and α2 in vertebrates (Supplementary Fig. 10) and which we have called the α-hook structure (Fig. 3b, d). The α-hook interacts with the γ-subunit at the exchangeable binding site 3, with AMP bound, which we have assigned as the weaker of the two sites responsible for mediating AMP/ADP protection against dephosphorylation. The hook makes a lid over the nucleotide-binding site that accounts for a buried surface area of about 500 Å[2]. We obtained crystals of this construct with AMP added to the crystallization mixture, and an AMP molecule is clearly identifiable in the electron density maps at site 3, as well as at the non-exchangeable site. Although we did not see AMP at site 1 in the initial crystals, we subsequently achieved good occupancy at this site by maintaining a higher concentration of AMP during crystallization and handling procedures (data not shown). In contrast, we think that the AMP at site 3 is held in place by the arrangement of the α-hook and that it would have to dissociate before the bound AMP could be released and then exchanged. On the basis of superposition of our earlier structures of the regulatory fragment of AMPK in complex with ADP and Mg-ATP and the structure presented here, we think that the α-hook sequence cannot interact with site 3 when Mg-ATP is bound, mainly
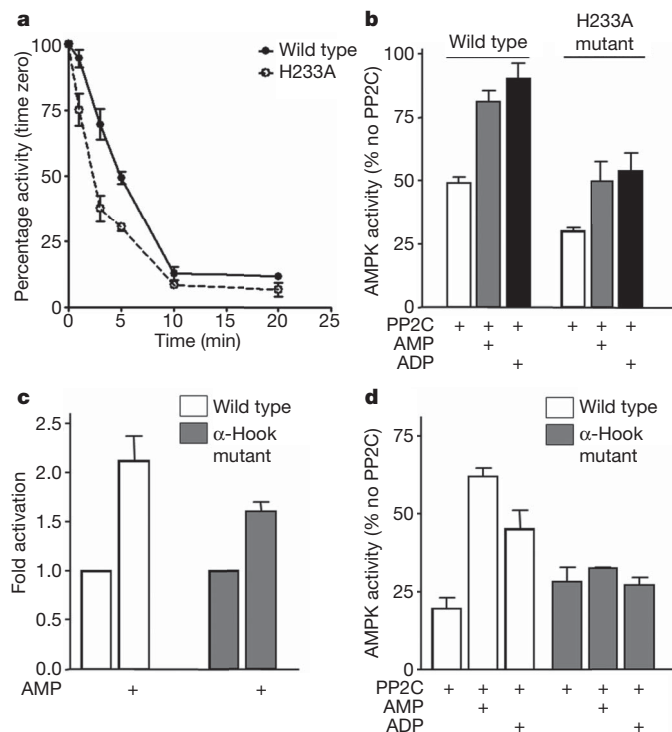

Figure 4 | Mutational analysis of AMPK regulation. a, Dephosphorylation rate of the wild type or β-His 233 to alanine kinase domain interface mutant (H233A, corresponding to H235A in β2). **b,** Protection of dephosphorylation of wild type or H233A mutant by AMP (30 μM) or ADP (30 μM) after incubation for 5 min. **c,** Allosteric activation of wild type or α-hook mutant (harbouring mutation of residues R375Q, T377A, D379A and E380A within α1) by AMP (100 μM). **d,** AMP (30 μM) and ADP (30 μM) protection of wild type or α-hook mutant from dephosphorylation after incubation for 20 min. All results are the mean ± s.e.m. from at least three independent experiments.

due to the change in conformation of Arg 69 of the γ-subunit, which would generate a steric clash with the α-hook (Supplementary Fig. 11). To test the role of the α-hook in mediating protection against dephosphorylation we generated a mutant in this region (R375Q/T377A/D379A/E380A). The resulting enzyme was allosterically activated by AMP but was not subject to protection against dephosphorylation by AMP or ADP (Fig. 4c, d). Interestingly, the mutation at His 233 described above does retain some protective effect of AMP/ADP (Fig. 4a, b). Given that this mutation would be expected to weaken the interaction between the kinase domain and the regulatory fragment, but not block it, it seems reasonable that AMP/ADP binding would still help to order the α-hook and thus facilitate the recruitment of the kinase domain.

Taking our biochemical and structural data together, we propose the following model for how AMP/ADP, but not Mg-ATP, protects AMPK against dephosphorylation, and thus inactivation (Supplementary Fig. 1). We have demonstrated that the protective effect of AMP/ADP is mediated by its binding to the weaker of the two exchangeable sites, which we have identified as site 3. We have also shown that the α-hook region binds into this site in the presence of AMP and predict that the same situation would occur with ADP. We further suggest that binding of the α-hook acts to restrict the flexibility of the preceding α-linker region (residues 300–370) and, in so doing, promotes the interaction of the kinase domain with the regulatory fragment seen in our crystal structure. This interaction, which mostly involves contacts between the activation loop and the C-terminal domain of β, would therefore act to protect Thr 172 against dephosphorylation. Because the interaction surface of the α-hook with the regulatory fragment is relatively small it is likely that there is a dynamic equilibrium between the α-hook-bound and α-hook-unbound species. If, as our structure indicates, AMP/ADP binding favours the α-hook-bound species but Mg-ATP binding drives formation of the α-hook-unbound species, then the competitive binding of AMP/ADP versus Mg-ATP would control the extent to which the enzyme was protected from dephosphorylation and inactivation.

METHODS SUMMARY

AMPK complexes were expressed in *Escherichia coli* BL21 (DE3) cells, purified by affinity chromatography using nickel-Sepharose and phosphorylated by incubation with CaMKKβ as described previously[9]. AMPK activity was determined using 0.2 mM SAMS peptide[9], 0.2 mM ATP and 5 mM MgCl₂. Dephosphorylation of AMPK by recombinant PP2C-α was monitored either by measuring AMPK activity using the SAMS peptide assay or by western blotting of phospho-T172. Western blot signals for phospho-T172 and total AMPK α-subunit (determined using sheep anti-α1 or anti-α2 antibodies) were quantified using the Li-Cor Odyssey infrared imaging system. Uncorrected fluorescence spectra of the nucleotides (3′-(7-diethylaminocoumarin-3-carbonylamino)-3′-deoxy-ADP (C-ADP), and 3′-(7-diethylaminocoumarin-3-carbonylamino)-3′-deoxy-ATP (C-ATP) (both gifts from M. Webb)) and NADH and their complexes were recorded at 20 °C using a Jasco FP-6300 fluorimeter. Binding of nucleotides was monitored by titrating nucleotide (4–10 μM) with AMPK. Dissociation constants for AMP, ADP and ATP were determined using competition assays. The engineered crystallization construct was expressed as a His-tag fusion protein in *E. coli*. Purified protein was phosphorylated using CAMKKβ before mixing with AMP and staurosporine. Crystals were grown by the hanging-drop method using isopropanol and MPD as precipitant. Diffraction data were collected on the Diamond Light Source, Oxford and processed using Denzo and Scalepack[28]. The structure was solved by molecular replacement using Amore[29] and standard refinement was carried out with Refmac5[30] with manual model building with COOT. Figures were created with Pymol (http://pymol.sourceforge.net/).

1. Carling, D. The AMP-activated protein kinase cascade–a unifying system for energy control. *Trends Biochem. Sci.* **29,** 18–24 (2004).
2. Hardie, D. G. AMP-activated/SNF1 protein kinases: conserved guardians of cellular energy. *Nature Rev. Mol. Cell Biol.* **8,** 774–785 (2007).
3. Kahn, B. B., Alquier, T., Carling, D. & Hardie, D. G. AMP-activated protein kinase: ancient energy gauge provides clues to modern understanding of metabolism. *Cell Metab.* **1,** 15–25 (2005).
4. Shackelford, D. B. & Shaw, R. J. The LKB1-AMPK pathway: metabolism and growth control in tumour suppression. *Nature Rev. Cancer* **9,** 563–575 (2009).
5. Cool, B. *et al.* Identification and characterization of a small molecule AMPK activator that treats key components of type 2 diabetes and the metabolic syndrome. *Cell Metab.* **3,** 403–416 (2006).
6. Huang, X. *et al.* Important role of the LKB1-AMPK pathway in suppressing tumorigenesis in PTEN-deficient mice. *Biochem. J.* **412,** 211–221 (2008).
7. Stein, S. C., Woods, A., Jones, N. A., Davison, M. D. & Carling, D. The regulation of AMP-activated protein kinase by phosphorylation. *Biochem. J.* **345,** 437–443 (2000).
8. Oakhill, J. S. *et al.* β-Subunit myristoylation is the gatekeeper for initiating metabolic stress sensing by AMP-activated protein kinase (AMPK). *Proc. Natl Acad. Sci. USA* **107,** 19237–19241 (2010).
9. Sanders, M. J., Grondin, P. O., Hegarty, B. D., Snowden, M. A. & Carling, D. Investigating the mechanism for AMP activation of the AMP-activated protein kinase cascade. *Biochem. J.* **403,** 139–148 (2007).
10. Minokoshi, Y. *et al.* Leptin stimulates fatty-acid oxidation by activating AMP-activated protein kinase. *Nature* **415,** 339–343 (2002).
11. Yamauchi, T. *et al.* Adiponectin stimulates glucose utilization and fatty-acid oxidation by activating AMP-activated protein kinase. *Nature Med.* **8,** 1288–1295 (2002).
12. Watt, M. J. *et al.* CNTF reverses obesity-induced insulin resistance by activating skeletal muscle AMPK. *Nature Med.* **12,** 541–548 (2006).
13. Andersson, U. *et al.* AMP-activated protein kinase plays a role in the control of food intake. *J. Biol. Chem.* **279,** 12005–12008 (2004).
14. Minokoshi, Y. *et al.* AMP-kinase regulates food intake by responding to hormonal and nutrient signals in the hypothalamus. *Nature* **428,** 569–574 (2004).
15. Hardie, D. G., Carling, D. & Sim, A. T. R. The AMP-activated protein kinase: a multisubstrate regulator of lipid metabolism. *Trends Biochem. Sci.* **14,** 20–23 (1989).
16. Suter, M. *et al.* Dissecting the role of 5′-AMP for allosteric stimulation, activation, and deactivation of AMP-activated protein kinase. *J. Biol. Chem.* **281,** 32207–32216 (2006).
17. Carling, D., Sanders, M. J. & Woods, A. The regulation of AMP-activated protein kinase by upstream kinases. *Int. J. Obes.* **32** (Suppl. 4), S55–S59 (2008).
18. Momcilovic, M., Hong, S. P. & Carlson, M. Mammalian TAK1 activates Snf1 protein kinase in yeast and phosphorylates AMP-activated protein kinase *in vitro. J. Biol. Chem.* **281,** 25336–25343 (2006).
19. Davies, S. P., Helps, N. R., Cohen, P. T. & Hardie, D. G. 5′-AMP inhibits dephosphorylation, as well as promoting phosphorylation, of the AMP-activated protein kinase. Studies using bacterially expressed human protein phosphatase-2Cα and native bovine protein phosphatase-2AC. *FEBS Lett.* **377,** 421–425 (1995).
20. Cheung, P. C., Salt, I. P., Davies, S. P., Hardie, D. G. & Carling, D. Characterization of AMP-activated protein kinase γ-subunit isoforms and their role in AMP binding. *Biochem. J.* **346,** 659–669 (2000).
21. Xiao, B. *et al.* Structural basis for AMP binding to mammalian AMP-activated protein kinase. *Nature* **449,** 496–500 (2007).
22. Kemp, B. E., Oakhill, J. S. & Scott, J. W. AMPK structure and regulation from three angles. *Structure* **15,** 1161–1163 (2007).
23. Carling, D., Clarke, P. R., Zammit, V. A. & Hardie, D. G. Purification and characterization of the AMP-activated protein kinase. Copurification of acetyl-CoA carboxylase kinase and 3-hydroxy-3-methylglutaryl-CoA reductase kinase activities. *Eur. J. Biochem.* **186,** 129–136 (1989).
24. Veech, R. L., Lawson, J. W., Cornell, N. W. & Krebs, H. A. Cytosolic phosphorylation potential. *J. Biol. Chem.* **254,** 6538–6547 (1979).
25. Hellsten, Y., Richter, E. A., Kiens, B. & Bangsbo, J. AMP deamination and purine exchange in human skeletal muscle during and after intense exercise. *J. Physiol.* **520,** 909–920 (1999).
26. McConell, G. K. *et al.* Short-term exercise training in humans reduces AMPK signalling during prolonged exercise independent of muscle glycogen. *J. Physiol. (Lond.)* **568,** 665–676 (2005).
27. Littler, D. R. *et al.* A conserved mechanism of autoinhibition for the AMPK kinase domain: ATP-binding site and catalytic loop refolding as a means of regulation. *Acta Crystallogr. F* **66,** 143–151 (2010).
28. Otwinowski, Z. & Minor, W. Data Collection and Processing. *Proc. CCP4 Study Weekend* 556–562 (SERC, 1993).
29. Navaza, J. AMoRe: an Automated Package for Molecular Replacement. *Acta Crystallogr. A* **50,** 157–163 (1994).
30. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).

Author Contributions B.X., M.J.S., E.U., R.H., F.V.M., D. Carmena, C.J., P.A.W., J.F.E., L.F.H., P.S., S.A.H., R.A. and S.R.M. performed experiments. All authors contributed to data analysis, experimental design and manuscript writing.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 2Y8L, 2Y8Q, 2Y94 and 2YA3. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.C. (david.carling@csc.mrc.ac.uk) or S.J.G. (sgambli@nimr.mrc.ac.uk).

LETTER

Structural basis for recognition of centromere histone variant CenH3 by the chaperone Scm3

Zheng Zhou[1]*, Hanqiao Feng[1]*, Bing-Rui Zhou[1], Rodolfo Ghirlando[2], Kaifeng Hu[3], Adam Zwolak[4], Lisa M. Miller Jenkins[5], Hua Xiao[1], Nico Tjandra[4], Carl Wu[1] & Yawen Bai[1]

The centromere is a unique chromosomal locus that ensures accurate segregation of chromosomes during cell division by directing the assembly of a multiprotein complex, the kinetochore[1]. The centromere is marked by a conserved variant of conventional histone H3 termed CenH3 or CENP-A (ref. 2). A conserved motif of CenH3, the CATD, defined by loop 1 and helix 2 of the histone fold, is necessary and sufficient for specifying centromere functions of CenH3 (refs 3, 4). The structural basis of this specification is of particular interest. Yeast Scm3 and human HJURP are conserved non-histone proteins that interact physically with the (CenH3–H4)$_2$ heterotetramer and are required for the deposition of CenH3 at centromeres in vivo[5–13]. Here we have elucidated the structural basis for recognition of budding yeast (Saccharomyces cerevisiae) CenH3 (called Cse4) by Scm3. We solved the structure of the Cse4-binding domain (CBD) of Scm3 in complex with Cse4 and H4 in a single chain model. An α-helix and an irregular loop at the conserved amino terminus and a shorter α-helix at the carboxy terminus of Scm3(CBD) wraps around the Cse4–H4 dimer. Four Cse4-specific residues in the N-terminal region of helix 2 are sufficient for specific recognition by conserved and functionally important residues in the N-terminal helix of Scm3 through formation of a hydrophobic cluster. Scm3(CBD) induces major conformational changes and sterically occludes DNA-binding sites in the structure of Cse4 and H4. These findings have implications for the assembly and architecture of the centromeric nucleosome.

Unlike other eukaryotic species that have complex regional centromeres with multiple centromeric nucleosomes[14], budding yeast has a single centromeric nucleosome that is necessary and sufficient to mediate the accurate segregation of chromosomes during mitosis and meiosis[15–18]. The simple centromeres of budding yeast provide an attractive system for investigating outstanding topics in centromere biology, including the pathway of CenH3 deposition and the architecture of the centromeric nucleosome[19,20].

Yeast Scm3 and human HJURP are binding partners of CenH3–H4 and are functionally required for their deposition at centromeres in vivo[5–13]. A conserved domain of Scm3 dictates specific and stoichiometric binding of CenH3–H4 (Fig. 1a), forming a (Scm3–Cse4–H4)$_2$ hexamer in 2 M NaCl[5]. This CBD of Scm3 is mapped to residues 84–187 (ref. 5). To investigate the structural basis for the recognition of Cse4 by Scm3, we first analysed the CBD of Scm3 by NMR and found that it is intrinsically disordered (Supplementary Fig. 1). To overcome instability inherent in complexes of individual Scm3, Cse4 and H4 fragments (Supplementary Figs 2–4), we engineered a single-chain molecule in which Scm3 is inserted between Cse4 and H4 to assemble a stably folded molecule (Supplementary Fig. 5). For convenience, we termed the single-chain molecule scSCH (Scm3, Cse4, H4).

The structure of scSCH was determined using multidimensional NMR and verified by structural analysis of its mutants (Fig. 1b–d and Supplementary Fig. 6). The structure of the folded core of scSCH, which includes residues 97–135 of Scm3, 157–202 of Cse4 and 50–99 of H4, is
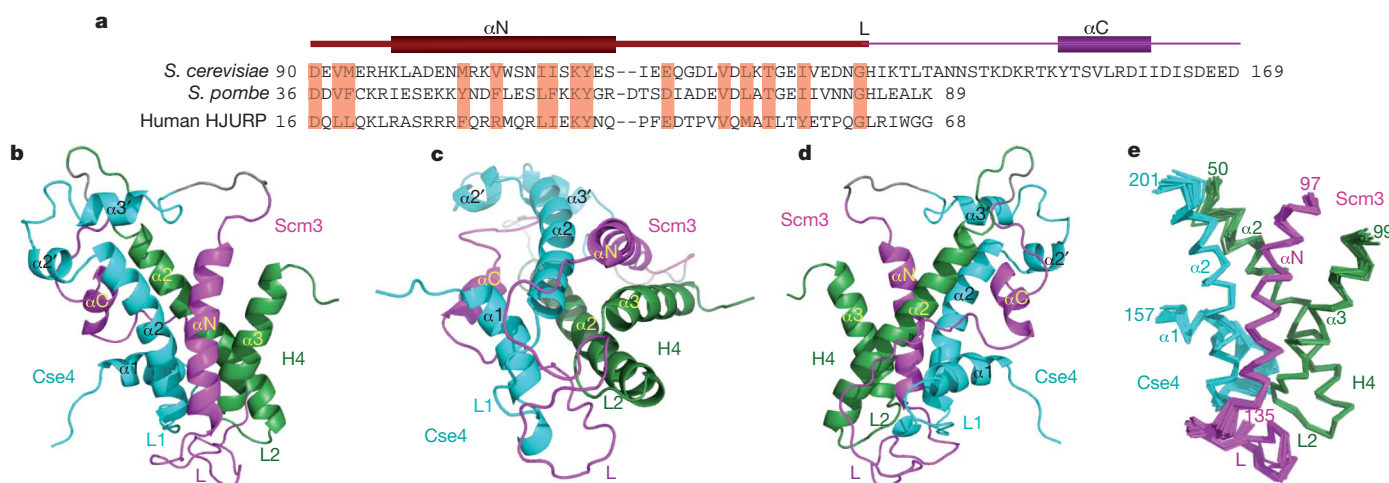

Figure 1 | Overall structure of scSCH. a, The amino acid sequence and secondary structures of the Cse4-binding domain of Scm3 in scSCH. Also shown are the conserved regions in Scm3 of S. pombe and human HJURP. Highly conserved residues are highlighted in red. The region in the folded core is shown in dark magenta (see **e**). **b–d**, Front, bottom and back view of the scSCH structure shown in ribbon representation. Scm3, Cse4 and H4 are in magenta, cyan and dark green, respectively. The full sequence of scSCH is M-His$_6$-KK-Cse4(150–227)-LVPRGS-Scm3(93–169)-GDK-H4(42–103).

[1]Laboratory of Biochemistry and Molecular Biology, National Cancer Institute, Bethesda, Maryland 20892, USA. [2]Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland 20892, USA. [3]National Magnetic Resonance Facility at Madison, University of Wisconsin, Madison, Wisconsin 53706, USA. [4]Laboratory of Molecular Biophysics, National Heart, Lung, and Blood Institute, NIH, Bethesda, Maryland 20892, USA. [5]Laboratory of Cell Biology, National Cancer Institute, NIH, Bethesda, Maryland 20892, USA.
*These authors contributed equally to this work.

well defined with root mean squared deviations (r.m.s.d.) of 0.54 Å for backbone atoms and 1.06 Å for all heavy atoms (Fig. 1e and Supplementary Table 1). Importantly, linker residues inserted between Scm3, Cse4 and H4 do not alter the structure of the folded region. Proteolytic cleavage of the two linkers in the folded scSCH only affects chemical shifts of neighbouring residues close to the cutting sites (Supplementary Fig. 7). Moreover, the folded structure of the above tertiary complex is unchanged by refolding after denaturation in 6 M GdmCl to liberate the three components as individual polypeptides (Supplementary Fig. 8). Backbone amide ^{15}N-{^{1}H} heteronuclear Overhauser effects (NOE) reflect dynamic motions. The folded core shows small dynamic motions (NOE >0.7) except for loop 1 of Cse4 and the small loop region following the N-terminal α-helix (αN) in Scm3 (Supplementary Fig. 9a, b). In contrast, other regions display larger dynamic motions (NOE <0.7), corresponding to less-well-defined structures (Supplementary Figs 9–11).

In the structure of scSCH, Scm3 interacts broadly with both Cse4 and H4. The αN helix of Scm3 makes close contacts with both the α2 helix of Cse4 and the α3 helix of H4 through multiple hydrophobic interactions (Figs 1b and 2a–d and Supplementary Figs 9c and 13a). Following the αN helix, the loop region of Scm3 (residues 121–144) mainly interacts with loop 1 of Cse4, except that a bulge (Scm3 residues 125–130) in the middle of the loop lies on top of loop 2 of H4 (Fig. 1c and Supplementary Figs 9d and 13b–d). Scm3 loop residues 140–144 also interact with the C-terminal portion of the α2 helix of H4 (Supplementary Figs 9d and 13e). Interestingly, Scm3 residues 145–154 are completely disordered (Supplementary Figs 9b, e and 10). Finally, the C-terminal α-helix (αC) of Scm3 (155–161) makes interactions with the N-terminal region of the α2 helix of H4 (Supplementary Fig. 13f).

Next, we analysed the effects of mutations on the formation of Scm3–Cse4–H4 complexes with isothermal titration calorimetry. The results reveal that the Scm3 recognition motif resides in the N-terminal region (181–190) of the α2 helix of Cse4. Double mutations Met181Ser/Met184Gly and Ala189Ser/Ser190Val in Cse4 that change the Cse4-specific residues to the corresponding residues in H3 reduced the binding affinity by a factor of 5.5 and 9, respectively (Fig. 2a, b, Supplementary Table 2 and Supplementary Fig. 14). A double mutation Ile110Asp/Ile117Asn in the αN helix of Scm3 decreased the binding affinity by a factor of 85 (Fig. 2a, Supplementary Fig. 14 and Supplementary Table 2). These residues are important for cell growth: mutation of the three residues (Met184, Ala189, and Ser190) in Cse4 to corresponding residues in H3 leads to growth defect (small colony)[21], and mutation Ile110Asp/Ile117Asn in Scm3 abrogates cell viability[7], consistent with the effects of these mutations on the binding affinity between Scm3 and Cse4/H4 (Supplementary Table 2). Met 181 should also be important for cell function because it interacts with Ile 117 of Scm3 (Fig. 2a). It is possible that simultaneous mutation of the four residues in Cse4 to the corresponding residues in H3 would abrogate cell viability.

In contrast, deleting the three extra residues Lys 172, Asp 173 and Gln 174 and mutating Thr 170 in loop 1 (to Lys, as in H3), all residues specific to Cse4 (Supplementary Fig. 12), had little effect on the binding affinity (a factor of 1.1) (Supplementary Table 2). Mutation of four residues (Val 165, Thr 166, Asp 167, Glu 168) at the C-terminal region of the α1 helix of Cse4 to corresponding residues in H3 (Ile, Ala, Gln and Asp) also showed little effect on the binding affinity (a factor of 1.4) (Supplementary Table 2). In addition, we found that Scm3 is capable of pulling down the H3^{CATD}–H4 chimaera, in which the CATD of Cse4 is swapped to the corresponding region of H3 (ref. 22 (Supplementary Fig. 15). Furthermore, Scm3 can pull down an H3 mutant with only four residues replaced by the corresponding residues in the α2 helix of Cse4 (Met 181, Met 184, Ala 188 and Ser 189) as well (Fig. 2c, d). Importantly, Scm3 residues that interact with the four Cse4-specific residues are well conserved in human HJURP (Fig. 1a). Indeed, Cse4 can also pull down the N-terminal region (residues 2–81) of HJURP (homologous to the Cse4-binding motif of Scm3 (refs 9, 22)) and human H4 (Supplementary Fig. 16). This result is consistent with the ability of Cse4 to replace human CENP-A at centromeres and maintain centromere function in human cells[23]. In addition, the CENP-A residues that correspond to the four Cse4-specific residues are reasonably conserved (Fig. 2d).

The structure of scSCH reveals the induction of major local conformational changes in the structure of Cse4 and H4 relative to the (CENP-A–H4)$_2$ tetramer. First, the packing between the central α2 helices of Cse4 and H4 in the structure of scSCH is loose in comparison to tight hydrophobic interactions in the homology-modelled dimer based on the H3–H4 structure in the nucleosome (Fig. 3a, b), or in the CENP-A–H4 dimer in the (CENP-A–H4)$_2$ tetramer[24]. Hydrophobic residues Leu 59, Phe 62 and Val 66 in the α2 helix of H4 lose interacting partners Tyr 193, Ser 192 and Leu 186 in the α2 helix of Cse4 (Fig. 3a, b; Phe 101, Ala 98 and Leu 94 in human CENP-A (ref. 24)). Second, owing to the insertion of the Scm3 loop, loop 1 in Cse4 loses close contact with loop 2 of H4 (Supplementary Fig. 13b) when compared with the corresponding loops in the canonical histone octamer[25] (Fig. 3c, d) or in the human (CENP-A–H4)$_2$ tetramer[24] (Supplementary Fig. 17). Third, the α2 helix of Cse4 kinks in the middle in scSCH (Fig. 4a), as forced by the side chain of residue Met 103 of the αN helix of Scm3 (Fig. 4a). In contrast, the α2 helix of CENP-A or H3 is relatively straight in the CENP-A–H4 (ref. 24) (Fig. 4b) or H3–H4 tetramer[25] (Supplementary Fig. 18). Fourth, the C-terminal region (94–99) of H4 in scSCH adopts a striking helical conformation and extends the α3 helix of H4. The helical conformation is induced by the side chain
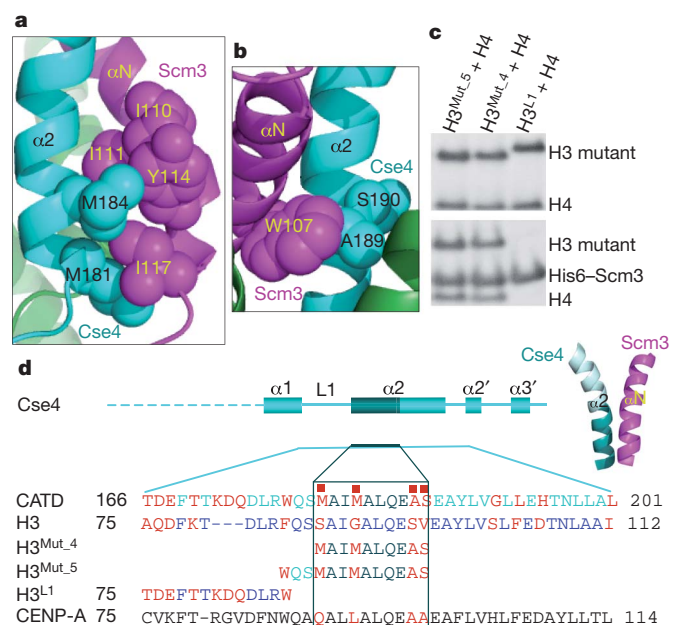

Figure 2 | The N-terminal region (181–190) of the α2 helix of Cse4 is the Scm3 recognition motif. a, Ile 110, Ile 111, Tyr 114 and Ile 117 (balls in magenta) in the αN helix of Scm3 form a hydrophobic cluster with Cse4-specific residues Met 181 and Met 184 (balls in cyan) in the α2 helix of Cse4. **b**, Trp 107 (balls in magenta) in the αN helix of Scm3 has close interactions with the Cse4-specific residue Ala 189 (balls in cyan) in the α2 helix of Cse4. Ser 190 is also a Cse4-specific residue (balls in cyan). **c**, SDS–PAGE gels showing the pull-down results with mutants of H3. The top panel shows the input of H3 mutants and H4. The bottom panel shows the molecules eluted from His$_6$–Scm3 (Scm3(65–169))-bound beads with 250 mM imidazole. H3$^{Mut\_4}$, H3$^{Mut\_5}$ and H3^{L1} are the mutants of H3 (see **d**). **d**, Illustration of the secondary structures in Cse4 and the Scm3 recognition motif (dark cyan), CATD, and the mutants used in the pull-down experiments. The red squares indicate the four residues that are sufficient for specific recognition of Cse4 by Scm3. The sequences swapped from Cse4 to H3 in the mutations are shown. The sequences that are not changed in the swap are omitted.
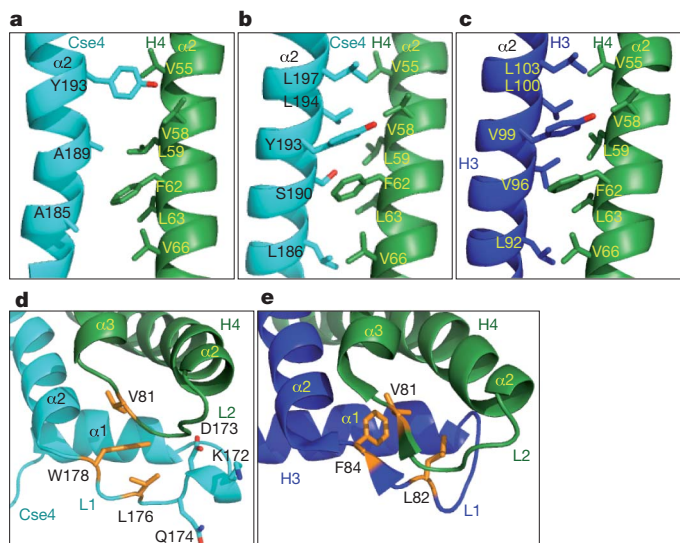
Figure 3 | Altered interactions in the CATD region in scSCH. a, The region of the α2 helices of Cse4 and H4 in scSCH, showing that there is little side-chain interaction between the two helices. **b**, The corresponding region of the α2 helices of Cse4 and H4 in the Cse4–H4 dimer structure obtained by homology modelling based on the structures of H3 and H4 in the nucleosome, showing that there are many hydrophobic interactions. **c**, The region of the α2 helices of H3 and H4 in the nucleosome. **d**, Region of loop 1 of Cse4 and the loop 2 of H4. The side chains of the hydrophobic residues are shown in stick representation and orange. The extra three residues in loop 1 of Cse4 are shown in stick representation. **e**, The corresponding loop 1 of H3 and loop 2 of H4 in the nucleosome structure. The hydrophobic residues are shown in stick representation and orange.

of Leu 98 in the αN helix of Scm3 through hydrophobic interactions with the side chains of Leu 98 and Tyr 99 of H4 (Fig. 4a, Supplementary Fig. 14 and Supplementary Table 2). This region is presumably disordered in the (CENP-A–H4)₂ tetramer[24] (Fig. 4b). Interestingly, the same region forms a mini β-strand that pairs with a β-strand of H2A in the canonical histone octamer or with a β-strand of histone
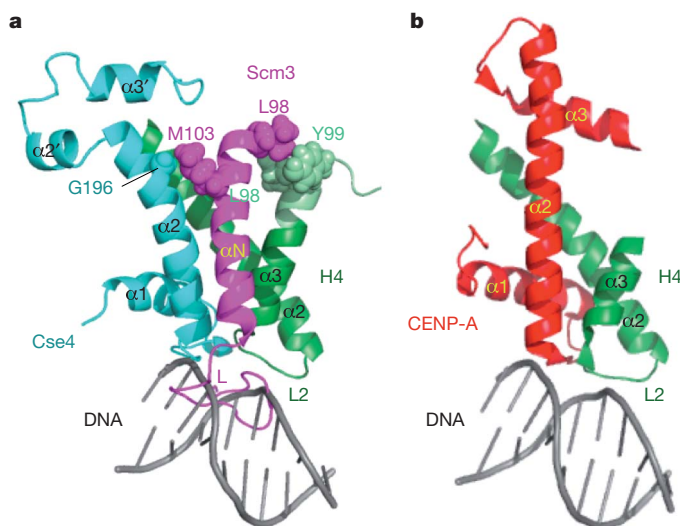

Figure 4 | Scm3 induces large conformational changes in Cse4 and H4 and prevents loop 2 of H4 from binding to DNA. a, Cse4–H4 in scSCH. The extended α3 helix in H4 is shown in light green. The loop of Scm3 pushes loop 1 of Cse4 away from loop 2 of H4 and prevents loop 2 of H4 from binding to DNA. DNA is modelled to bind the loop 2 region of H4 based on the canonical nucleosome structure[29]. **b**, CENP-A–H4 in the (CENP-A–H4)₂ tetramer. DNA is modelled to bind to the loop 2 region of H4 based on the canonical nucleosome struture[29].

chaperone Asf1 in the Asf1–H3–H4 complex[26,27] (Supplementary Fig. 18).

Furthermore, the C-terminal region of Cse4 has considerable disorder in scSCH (Figs 4a, b). The same region is also disordered in the Cse4–H4 dimer and is not required for Cse4–H4 binding to Scm3 (Supplementary Fig. 5 and Supplementary Fig. 19). This 'tetramerization domain' is well folded in the (CENP-A–H4)₂ tetramer. Structure modelling shows that imposing this folded domain on the corresponding region of Cse4 in scSCH allows association as a (Scm3–Cse4–H4)₂ hexamer without major structural incompatibility (Supplementary Fig. 20), consistent with the existence of (Scm3–Cse4–H4)₂ hexamers in 2 M NaCl (ref. 5). However, in this context, histone topography in the scSCH structure outside the tetramerization domain displays dramatic global conformational changes when compared with the (CENP-A–H4)₂ tetramer, making the modelled (Scm3–Cse4–H4)₂ hexamer incompatible with DNA binding (Supplementary Fig. 20). Moreover, the Scm3 loop in the scSCH structure blocks loop 2 of H4, which makes contacts with DNA in the canonical nucleosome (Fig. 4a, b).

Thus, the structure of scSCH indicates that retention of Scm3 in association with centromere DNA is unlikely to occur via binding of Scm3(CBD) to Cse4/H4, as binding of DNA and Scm3(CBD) to Cse4/H4 is mutually incompatible. Instead, Scm3(CBD) behaves as a specific histone chaperone, and the retention of Scm3 with Cse4/H4 on centromeric DNA requires its distinct DNA-binding domain (H. Xiao and C. Wu, manuscript in preparation).

The structure of scSCH reveals that a subregion within the CATD, including four Cse4-specific residues in the N-terminal region of the α2 helix of Cse4, is necessary and sufficient for specific recognition by Scm3. Thus, the remainder of the CATD of Cse4 should be important for association with other proteins for Cse4 functions. The CBD of Scm3 uses both induced histone conformation changes[26,27] and direct steric occlusion[28] to prevent Cse4–H4 in the Scm3–Cse4–H4 complex from DNA binding (Supplementary Fig. 21). Conversely, Cse4–H4, with a conformation similar to that of CENP-A–H4 in the (CENP-A–H4)₂ tetramer, is unfavourable for Scm3(CBD) binding but favours DNA binding, indicating a competition mechanism for Scm3 and HJURP as CenH3-specific chaperones[22].

METHODS SUMMARY

1. Cleveland, D. W., Mao, Y. & Sullivan, K. F. Centromeres and kinetochores: From epigenetics to mitotic checkpoint signaling. *Cell* **112**, 407–421 (2003).
2. Henikoff, S., Ahmad, K. & Malik, H. S. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* **293**, 1098–1102 (2001).
3. Black, B. E. *et al.* Structural determinants for generating centromeric chromatin. *Nature* **430**, 578–582 (2004).
4. Black, B. E. *et al.* Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. *Mol. Cell* **25**, 309–322 (2007).
5. Mizuguchi, G., Xiao, H., Wisniewski, J., Smith, M. M. & Wu, C. Nonhistone Scm3 and histones CenH3–H4 assemble the core of centromere-specific nucleosomes. *Cell* **129**, 1153–1164 (2007).

6. Camahort, R. *et al.* Scm3 is essential to recruit the histone H3 variant Cse4 to centromeres and to maintain a functional kinetochore. *Mol. Cell* **26,** 853–865 (2007).
7. Stoler, S. *et al.* Scm3, an essential *Saccharomyces cerevisiae* centromere protein required for G2/M progression and Cse4 localization. *Proc. Natl Acad. Sci. USA* **104,** 10571–10576 (2007).
8. Camahort, R. *et al.* Cse4 is part of an octameric nucleosome in budding yeast. *Mol. Cell* **35,** 794–805 (2009).
9. Sanchez-Pulido, L., Pidoux, A. L., Pointing, C. P. & Allshire, R. C. Common ancestry of the CENP-A chaperones Scm3 and HJURP. *Cell* **137,** 1173–1174 (2009).
10. Williams, J. S., Hayashi, T., Yanagida, M. & Russell, P. Fission yeast Scm3 mediates stable assembly of Cnp1/CENP-A into centromeric chromatin. *Mol. Cell* **33,** 287–298 (2009).
11. Pidoux, A. L. *et al.* Fission yeast Scm3: A CENP-A receptor required for integrity of subkinetochore chromatin. *Mol. Cell* **33,** 299–311 (2009).
12. Foltz, D. R. *et al.* Centromere-specific assembly of CENP-A nucleosomes is mediated by HJURP. *Cell* **137,** 472–484 (2009).
13. Dunleavy, E. M. *et al.* HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell* **137,** 485–497 (2009).
14. Malik, H. S. & Henikoff, S. Major evolutionary transitions in centromere complexity. *Cell* **138,** 1067–1082 (2009).
15. Stoler, S., Keith, K. C., Curnick, K. E. & Fitzgerald-Hayes, M. A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. *Genes Dev.* **9,** 573–586 (1995).
16. Meluh, P. B., Yang, P., Glowczewski, L., Koshland, D. & Smith, M. M. Cse4p is a component of the core centromere of *Saccharomyces cerevisiae. Cell* **94,** 607–613 (1998).
17. Cottarel, G., Shero, J. H., Hieter, P. & Hegemann, J. H. A 125-base-pair CEN6 DNA fragment is sufficient for complete meiotic and mitotic centromere functions in *Saccharomyces cerevisiae. Mol. Cell. Biol.* **9,** 3342–3349 (1989).
18. Clarke, L. & Carbon, J. Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature* **287,** 504–509 (1980).
19. Black, B. E. & Bassett, E. A. The histone variant CENP-A and centromere specification. *Curr. Opin. Cell Biol.* **20,** 91–100 (2008).
20. Furuyama, T. & Henikoff, S. Centromeric nucleosomes induce positive DNA supercoils. *Cell* **138,** 104–113 (2009).
21. Keith, K. C. *et al.* Analysis of primary structural determinants that distinguish the centromere-specific function of histone variant Cse4p from histone H3. *Mol. Cell. Biol.* **19,** 6130–6139 (1999).
22. Shuaib, M., Ouararhni, K., Dimiyrov, S. & Hamiche, A. HJURP binds CENP-A via a highly conserved N-terminal domain and mediates its deposition at centromeres. *Proc. Natl Acad. Sci. USA* **107,** 1349–1354 (2010).
23. Wieland, G., Orthaus, S., Ohndorf, S., Diekmann, S. & Hemmerich, P. Functional complementation of human centromere protein A (CENP-A) by Cse4p from *Saccharomyces cerevisiae. Mol. Cell. Biol.* **24,** 6620–6630 (2004).
24. Sekulic, N., Bassett, E. A., Rogers, D. J. & Black, B. E. The structure of (CENP-A–H4)$_2$ reveals physical features that mark centromeres. *Nature* **467,** 347–351 (2010).
25. Wood, C. M. *et al.* High-resolution structure of the native histone octamer. *Acta Crystallogr.* **61,** 541–545 (2005).
26. English, C. M., Adkins, M. W., Carson, J. J., Churchill, M. E. & Tyler, J. K. Structural basis for the histone chaperone activity of Asf1. *Cell* **127,** 495–508 (2006).
27. Natsume, R. *et al.* Structure and function of the histone chaperone CIA/ASF1 complexed with histones H3 and H4. *Nature* **446,** 338–341 (2007).
28. Zhou, Z. *et al.* NMR structure of chaperone Chz1 complexed with histones H2A.Z-H2B. *Nature Struct. Mol. Biol.* **15,** 868–869 (2008).
29. White, C. L., Suto, R. K. & Luger, K. Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *EMBO J.* **20,** 5207–5218 (2001).

METHODS

Protein sample preparation. All proteins were expressed in *E. coli* (BL21-codonPlus(DE3)-RIL) with pET vectors (Stratagene). N-terminal His$_6$-tagged fragments of Cse4 and Scm3 and their mutants were first purified via Ni-NTA (Qiagen) whereas H4 and non-His-tagged fragments of Cse4 and Scm3 were first purified via SP sepharose (GE Healthcare). They were next subjected to reverse-phase HPLC purification using acetonitrile and water as solvents. Purified proteins were lyophilized. Isotope-labelled proteins for NMR studies were produced by growing *E. coli* cells in M9 media with ^{15}NH$_4$Cl, U-^{13}C$_6$-glucose, and D$_2$O as the sole source for nitrogen, carbon and deuterium, respectively. For the measurement of side-chain NOEs, specific methyl labelling (-^{13}CH$_3$) for Ile, Leu and Val residues was also made following the protocol of ref. 30.

To prepare the Cse4, H4 and Scm3 complexes, lyophilized proteins were first dissolved in H$_2$O. Their concentrations were determined by measuring the absorbance at 280 nm. Equal amounts of each species were mixed together and dialysed against 10 mM Tris-HCl and 2 M NaCl at pH 7.4 and 4 °C. After centrifugation, the soluble fractions were subjected to gel filtration on Superdex 75 10/300 GL column (GE Healthcare). The eluted complexes were concentrated with an Amicon with Ultra Ultracel-10 membrane (Millipore) and exchanged to a final buffer of 50 mM MES at pH 5.6. The Cse4–H4 complexes were made in the same way. Scm3(80–211) samples for NMR study were prepared by dissolving them in 8 M urea and dialysis against corresponding buffer. scSCH and all other single-chain molecules derived from scSCH are purified with Ni-NTA (Qiagen) under native conditions (20 mM Tris-HCl and 0.5 M NaCl at pH 8.0), followed by gel filtration with Superdex 200 10/60 column at 4 °C (GE healthcare). The fractions containing the target protein were combined and concentrated and exchanged with the final buffer (50 mM MES at pH 5.4).

Analytical ultracentrifugation. Sedimentation velocity experiments were conducted in duplicate at 20.0 °C on a Beckman Coulter Proteome XL-I analytical ultracentrifuge. 400 µl of the sample of 35 µM in 50 mM MES (pH 5.6) was loaded in a double sector centrepiece cell and analysed at a rotor speed of 50,000 r.p.m. One-hundred scans were acquired as single absorbance measurements ($\lambda = 280$ nm) at 7.1-min intervals using a radial spacing of 0.003 cm. Data were analysed in SEDFIT 11.71 in terms of a continuous $c(s)$ distribution to obtain a sedimentation coefficient, s, and molecular mass M (ref. 31). Solution densities ρ were measured at 20.0 °C on a Mettler Toledo DE51 density meter and solution viscosities η were measured using a Cannon–Ubbelohde viscometer and Cannon-CT 500 constant temperature bath set at 20.00 °C. The partial specific volume v of the complex was calculated in SEDNTERP 1.09 (ref. 32). $c(s)$ analyses were carried out using an s-value range of 0.5 to 6.0 with a linear resolution of 100 and a confidence level (F-ratio) of 0.68. The analyses, implemented using time-independent noise corrections, returned root mean square deviation (r.m.s.d.) values for the best fits of 0.0040 absorbance units. Sedimentation equilibrium experiments were conducted at 20.00 °C on a Beckman Optima XL-A. 135 µl volumes of the complex were studied at loading concentrations of 20, 39 and 78 µM, along with the sample recovered from the sedimentation velocity experiments. Experiments were carried out using six-channel centrepiece cells at rotor speeds ranging from 18,000 to 34,000 r.p.m. In all cases data were acquired as an average of four absorbance measurements at wavelengths of 280 and 250 nm using a radial spacing of 0.001 cm. Sedimentation equilibrium at each speed was achieved within 40 h. Data were analysed globally in terms of a single ideal species using SEDPHAT 6.21 (refs 32, 33).

NMR experiments. NMR experiments were performed on Bruker 500, 600, 800 and 900 MHz and Varian 600 and 800 MHz spectrometers at 35 °C. The following experiments were recorded. 2D: [^{1}H, ^{1}H]-NOESY, [^{1}H, ^{15}N]-TROSY, [^{1}H, ^{13}C]-HMQC, ^{15}N-{^{1}H} NOE; TROSY version 3D: HNCACB, HNCOCACB, HNCA, HNCOCA, HNCO, HNCACO; 3D HBHACONH, HCCH-TOCSY, CCH-TOCSY, CCC(CO)NH, [^{1}H,^{15}N]-NOESY-HSQC, [^{1}H,^{15}N]-NOESY-HSQC ([^{13}C]methyl-labelled sample), [^{1}H, ^{13}C]-NOESY-HSQC, [^{1}H,^{13}C]-NOESY-HSQC ([^{13}C]methyl-labelled sample). The spectra were processed using NMRPipe[34] and analysed with NMRView[35].

Structure calculation. Structure calculation was done using the program Xplor-NIH[36]. The NOE-derived restraints were subdivided into four classes, strong, medium, weak and very weak, by comparison with NOEs of protons separated by known distances as described previously[37]. Backbone dihedral angle restraints (ϕ and ψ angles) were obtained from analysis of ^{1}Hα, HN, ^{13}Cα, ^{13}Cβ, ^{13}CO and ^{15}N chemical shifts by using the program TALOS[38]. Two constraints per hydrogen bond (dNH-O ≤ 2.2 Å and dN-O ≤ 3.2 Å) were added in the final structure calculation after initial NOE-derived structures were obtained. The program PROCHECK_NMR[39] was used to evaluate the quality of the calculated structures.

Isothermal titration calorimetric experiments. The ITC experiments were performed on a MicroCal VP-ITC by injecting Scm3(83–169) solution (250 µM) to a solution of single-chain Cse4–H4 or their mutants (His$_6$-KK-Cse4(151–207)-LVPRGS-H4(45–103)) (20 µM) in a chamber of 1.4 ml at 25 °C in 50 mM MES (pH 5.4) and 0.1 M NaCl. Twenty-nine injections (each of 10 µl) were made and the heat released was analysed. The data were analysed as described previously[40].

Pull-down experiments. Pull-down experiments were carried out in 50 mM sodium phosphate, 25 mM imidazole, 2 M NaCl, pH 8.0 at room temperature. Ni-NTA (Qiagen) beads were mixed with His$_6$-Scm3(66–169) with a final concentration of 6 µM. Approximately 10-fold excess of (Cse4–H4)$_2$ or (H3–H4)$_2$ or their mutants was mixed with beads and incubated. The incubation was at 25 °C for 30 min. The beads were washed with the same buffer three times. The complex formed on the beads was eluted with 250 mM imidazole and analysed by SDS–PAGE. Beads without His$_6$-Scm3 were also incubated with corresponding (Cse4–H4)$_2$ under identical conditions to assess background binding and the integrity of the tetramer. No nonspecific binding was identified in 2 M NaCl. For molecules derived from thrombin-digested single-chain proteins, the complex was incubated with Ni-NTA (Qiagen) beads at 25 °C for 30 min and then washed three times. The final complex formed on the beads was eluted with either 8 M urea or 250 mM imidazole. The eluted molecules were analysed by SDS–PAGE.

30. Tugarinov, V., Kanelis, V. & Kay, L. E. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nature Protocols* **1**, 749–754 (2006).
31. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity untracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
32. Cole, J. L., Lary, J. W., Moody, T. P. & Laue, T. M. Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Methods Cell Biol.* **84**, 143–179 (2008).
33. Schuck, P. On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem.* **320**, 104–124 (2003).
34. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
35. Johnson, B. A. & Blevins, R. A. NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).
36. Schwieters, C. D., Kuszewski, J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73 (2003).
37. Zhou, Z. *et al.* NMR structure of chaperone Chz1 complexed with histones H2A.Z-H2B. *Nature Struct. Mol. Biol.* **15**, 868–869 (2008).
38. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999).
39. Laskowski, R. A. *et al.* AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).
40. Houtman, J. C. *et al.* Binding specificity of multiprotein signaling complexes is determined by both cooperative interactions and affinity preferences. *Biochemistry* **43**, 4170–4178 (2004).

LETTER

A conserved mechanism of DEAD–box ATPase activation by nucleoporins and InsP$_6$ in mRNA export

Ben Montpetit[1,2]*, Nathan D. Thomsen[2,3]†*, Kara J. Helmke[1,2], Markus A. Seeliger[2,3]†, James M. Berger[2,3] & Karsten Weis[1,2]

Superfamily 1 and superfamily 2 RNA helicases are ubiquitous messenger-RNA–protein complex (mRNP) remodelling enzymes that have critical roles in all aspects of RNA metabolism[1,2]. The superfamily 2 DEAD-box ATPase Dbp5 (human DDX19) functions in mRNA export and is thought to remodel mRNPs at the nuclear pore complex (NPC)[3–8]. Dbp5 is localized to the NPC via an interaction with Nup159 (NUP214 in vertebrates)[3–5,8,9] and is locally activated there by Gle1 together with the small-molecule inositol hexakisphosphate (InsP$_6$)[10,11]. Local activation of Dbp5 at the NPC by Gle1 is essential for mRNA export in vivo[10,12]; however, the mechanistic role of Dbp5 in mRNP export is poorly understood and it is not known how Gle1$_{InsP6}$ and Nup159 regulate the activity of Dbp5. Here we report, from yeast, structures of Dbp5 in complex with Gle1$_{InsP6}$, Nup159/Gle1$_{InsP6}$ and RNA. These structures reveal that InsP$_6$ functions as a small-molecule tether for the Gle1–Dbp5 interaction. Surprisingly, the Gle1$_{InsP6}$–Dbp5 complex is structurally similar to another DEAD-box ATPase complex essential for translation initiation, eIF4G–eIF4A, and we demonstrate that Gle1$_{InsP6}$ and eIF4G both activate their DEAD-box partner by stimulating RNA release. Furthermore, Gle1$_{InsP6}$ relieves Dbp5 autoregulation and cooperates with Nup159 in stabilizing an open Dbp5 intermediate that precludes RNA binding. These findings explain how Gle1$_{InsP6}$, Nup159 and Dbp5 collaborate in mRNA export and provide a general mechanism for DEAD-box ATPase regulation by Gle1/eIF4G-like activators.

Gle1 and InsP$_6$ robustly activate the ATPase activity of Dbp5 (refs 10, 11), but interactions between Dbp5 and Gle1$_{InsP6}$ are weak and transient in vitro[10,12]. However, gain-of-function mutations in Dbp5 and Gle1 (Dbp5(L327V) and Gle1(H337R))[10] allowed for the purification of a stable Dbp5–Gle1$_{InsP6}$ complex for protein crystallography (data not shown). Multiwavelength anomalous diffraction (MAD) phased protein crystal structures were initially obtained for Gle1$_{InsP6}$ and the carboxy-terminal RecA-like domain (C-terminal domain, CTD) of Dbp5(L327V), both in the presence and absence of the Gle1 H337R mutation at a resolution of 2.6 Å (R_{work}/R_{free} of 18.7/21.3%) and 2.5 Å (R_{work}/R_{free} of 20.2/23.8%), respectively (Supplementary Table 1). This model allowed us to solve the crystal structure of an amino-terminally truncated Δ90Dbp5(L327V) construct containing both RecA-like domains of Dbp5 (residues 91–482) bound to ADP and Gle1(H337R)$_{InsP6}$ (residues 244–538) by molecular replacement at a resolution of 4.0 Å (R_{work}/R_{free} of 21.4/23.9%) (Fig. 1 and Supplementary Table 2). Structures were also obtained for Δ90Dbp5(L327) or Δ90Dbp5 wild-type in complex with RNA and ADP•BeF$_3$ at a resolution of 1.5 Å (R_{work}/R_{free} of 16.5/18.9%) and 1.4 Å (R_{work}/R_{free} of 16.0/18.0%), respectively (Supplementary Table 2).

The structures reveal that Gle1 comprises an all α-helix HEAT repeat protein that makes contact with both RecA-like domains of

Dbp5 (Fig. 1a, b). The CTD interface buries 2,230 Å^2 of surface area, and contains a number of residues previously shown to be important for Gle1-mediated activation of Dbp5 (ref. 12). The second interaction surface uses the N-terminal RecA-like domain (N-terminal domain,
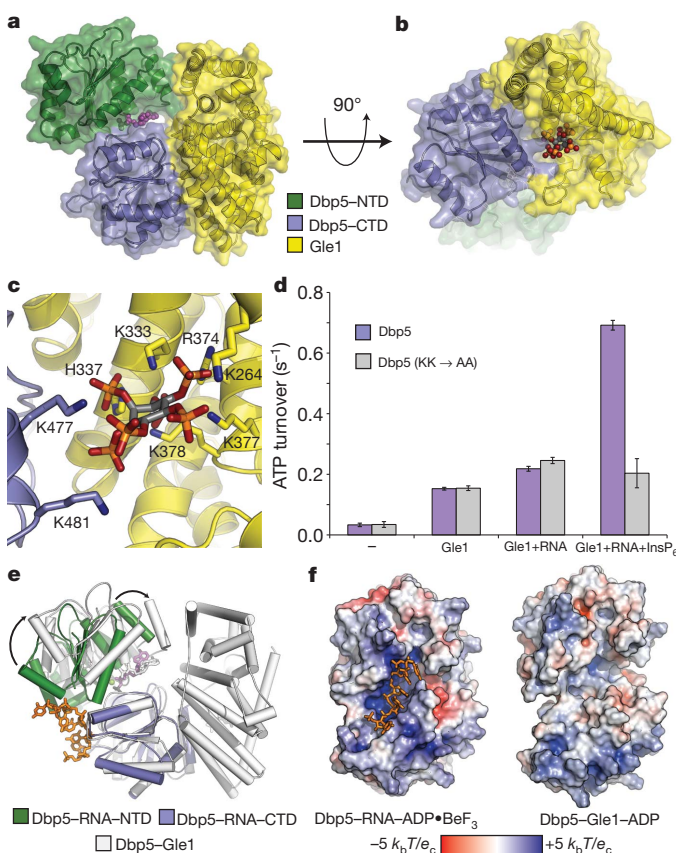

Figure 1 | The Gle1$_{InsP6}$–Δ90Dbp5–ADP complex. a, Structure of Dbp5 bound to ADP (magenta spheres) and Gle1 (see colour key). **b**, Side view showing InsP$_6$ (coloured spheres) bound at the interface between Gle1 and Dbp5. **c**, Detailed view of the InsP$_6$ binding interface (see colour key). InsP$_6$ is shown as grey sticks with orange phosphate and red oxygen atoms. Nitrogen atoms are in dark blue. **d**, InsP$_6$ binding is required for maximal Dbp5 ATPase stimulation with RNA. Error bars represent standard deviation (s.d.) ($n = 3$). **e**, Structural superposition of the RNA–Δ90Dbp5 complex with Gle1$_{InsP6}$–Δ90Dbp5. Arrows depict the large rigid body movement in Dbp5. **f**, Van der Waals surface view of the RNA–Δ90Dbp5 complex and the Gle1$_{InsP6}$–Δ90Dbp5 complex coloured by solvent accessible electrostatics. k_b, Boltzmann's constant (Joules/Kelvin); T, temperature (310 Kelvin); e_c, charge of an electron (1.602 × 10^{-19} Coulombs).

[1]Division of Cell and Developmental Biology, University of California, Berkeley, California 94720, USA. [2]Department of Molecular and Cell Biology and QB3 Institute, University of California, Berkeley, California 94720, USA. [3]Division of Biochemistry and Molecular Biology, University of California, Berkeley, California 94720, USA. †Present addresses: Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158, USA (N.D.T.); Department of Pharmacology, State University of New York at Stony Brook, Stony Brook, New York 11794, USA (M.A.S.).
*These authors contributed equally to this work.

NTD) of Dbp5 and an exposed hydrophobic α-helix of Gle1, which buries an additional 930 Å² of surface area (Fig. 1a). Interestingly, and despite the lack of obvious homology in sequence alignments, a search of the structural database using the DALI server[13] showed that the Gle1 fold is similar to that of eIF4G (14% identity, 4.1 Å root mean squared deviation (r.m.s.d.)), a co-factor for the superfamily 2 (SF2) DEAD-box ATPase eIF4A[14–16], and to that of Upf2 (also known as Nmd2; 8% identity, 3.0 Å r.m.s.d.), which stimulates the superfamily 1 (SF1) protein Upf1 (also known as Nam7)[17–19].

A unique feature of the Gle1$_{InsP6}$–Dbp5 structures is the presence of InsP$_6$, which bridges the protein–protein interaction by binding in a positively charged pocket at the interface between Gle1 and the CTD of Dbp5 (Fig. 1b, c). Residues K264, K333, H337, R374, K377 and K378 in Gle1 and K477 and K481 in Dbp5 interact with phosphate groups in InsP$_6$ (Fig. 1c and Supplementary Fig. 1a, b). Two of these residues in Gle1 (K377 and K378) were recently identified as being involved in InsP$_6$ binding[20]. Gle1 residues (H337, R374, K377 and K378) show charge conservation, and Dbp5 residue K477 is conserved as a hydrogen bond donor in most metazoans, indicating an evolutionary pressure to maintain a polar residue that can interact with InsP$_6$ at these positions. (Supplementary Fig. 1c, d). To our knowledge, this represents the first structural characterization of an endogenous small molecule that localizes at the interface of two heterologous proteins to mediate a protein–protein binding interaction.

To assess the importance of InsP$_6$ binding by Dbp5, we mutated the InsP$_6$ contacts K477 and K481 to Ala (Dbp5(KK→AA)). Dbp5(KK→AA) showed no synergistic ATPase activation in the presence of InsP$_6$, RNA and sub-saturating amounts of Gle1, but was still active as an ATPase and was stimulated by RNA and Gle1 comparable to wild-type protein in the absence of InsP$_6$ (Fig. 1d). Furthermore, Dbp5(KK→AA) (L327V variant) no longer formed a stable complex with Gle1(H337R) and InsP$_6$ (Supplementary Fig. 1e). These results demonstrate that InsP$_6$ stabilizes the Gle1–Dbp5 interaction by acting as a small-molecule tether.

A comparison of the RNA and ADP•BeF$_3$-bound Δ90Dbp5 structure with that of Gle1$_{InsP6}$–Dbp5 reveals that Gle1$_{InsP6}$ in the presence

of ADP induces a large rigid body movement in Dbp5, opening the tandem RecA-like domains relative to their positions in the RNA-bound state (Fig. 1e and Supplementary Movie 1). Smaller conformational changes occur at the Gle1$_{InsP6}$–CTD interface, and structures obtained with wild-type versus gain-of-function mutations show that these mutations subtly alter Gle1$_{InsP6}$–Dbp5 contacts, probably contributing additional binding energy to stabilize the complex (Supplementary Fig. 2). Interestingly, binding to Gle1$_{InsP6}$ markedly alters the RNA binding site in Dbp5, which opens and loses its localized positive charge (Fig. 1f and Supplementary Movie 2). This indicates that Gle1$_{InsP6}$ binding leads to the formation of a partially open, catalytically inactive and ADP-bound complex that precludes RNA binding.

Owing to the unexpected similarity between Gle1 and eIF4G, we also compared our structures to the eIF4A/4G complex. The open conformation of Dbp5 induced by Gle1$_{InsP6}$ is similar to that described for eIF4A in the presence of eIF4G[14,21,22], and there is an overall similarity between the interactions of eIF4G and Gle1$_{InsP6}$ with the CTDs of their DEAD-box ATPase partners (Fig. 2a). However, eIF4G lacks an N-terminal helix that in Gle1 forms part of the binding pocket for InsP$_6$; instead, eIF4G forms a flexible loop that makes additional contacts with eIF4A, effectively compensating for the lack of InsP$_6$ (Fig. 2b). The interface between eIF4G and the eIF4A NTD also exhibits interdomain differences when compared to Gle1$_{InsP6}$–Dbp5 (Fig. 2c, d), causing eIF4G to stabilize a more open eIF4A intermediate[14,21]. Thus, the eIF4G–eIF4A complex seems to be more competent for ADP exchange than the Gle1$_{InsP6}$–Dbp5 complex. Despite these small differences, both Gle1$_{InsP6}$ and eIF4G stabilize their DEAD-box binding partner in an open state that is incompatible with RNA binding.

Gle1$_{InsP6}$ is a potent activator of the overall ATPase activity of Dbp5 (refs 10, 11), but in our structures, Gle1 is not positioned to directly stimulate ATP hydrolysis (Fig. 1). We therefore proposed that Gle1$_{InsP6}$ activates Dbp5 by enhancing RNA release, a step known to be rate-limiting in the hydrolytic cycle of other DEAD-box ATPases[23,24]. In such a model, product release would rely on the ability of Gle1$_{InsP6}$ to induce the structural rearrangement of the two tandem RecA-like
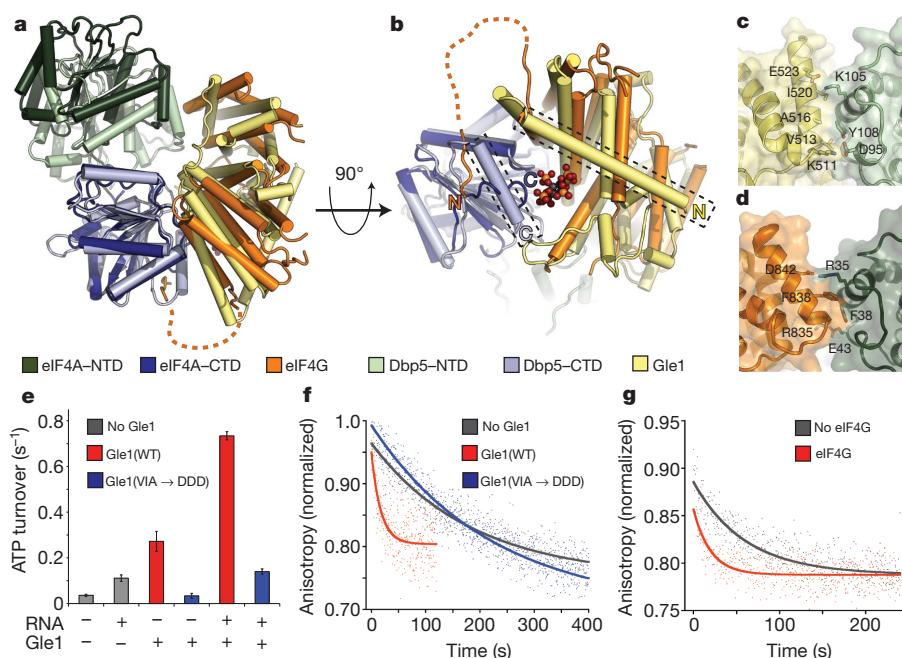
Figure 2 | Comparison of Gle1$_{InsP6}$–Δ90Dbp5 and eIF4A–eIF4G.
a, Structural superposition of Gle1$_{InsP6}$–Δ90Dbp5 with eIF4G–eIF4A and AMP (PDB accession code 2VSO)[21] (see colour key). **b**, View of the C-terminal RecA-like domain binding interface. Unique α-helices present in both Dbp5 and Gle1 form the InsP$_6$ biding pocket (boxed in figure). **c, d**, Residues involved in the

formation of the N-terminal RecA-like domain binding interface in Gle1$_{InsP6}$–Δ90Dbp5 (**c**) and eIF4A–eIF4G (**d**). **e**, Measured ATPase activity using wild-type (WT) or mutant Gle1. Error bars represent s.d. ($n = 3$). **f, g**, RNA release from Dbp5(E240Q) or eIF4A(E172Q) monitored by fluorescence polarization. Representative curves are shown.

domains of Dbp5 (Fig. 1e, f). Thus, loss of contact between Gle1 and the NTD of Dbp5 should impair the ability of Gle1$_{InsP6}$ to activate Dbp5 without affecting the binding between these proteins, which is primarily mediated by the Gle1$_{InsP6}$–CTD interaction (Fig. 1a). To test this idea, residues V513, A516 and I520 on a solvent-exposed helix of Gle1 (Supplementary Fig. 3a) were mutated (Gle1(VAI→DDD)) to disrupt the Gle1–NTD interface (Fig. 2c and Supplementary Fig. 3b). In an ATPase assay, Gle1(VAI→DDD) was unable to activate Dbp5 regardless of the presence of RNA and InsP$_6$ (Fig. 2e), but still formed a complex with Dbp5(L327V) by gel filtration (Supplementary Fig. 3c), demonstrating that these mutations did not disrupt the primary Gle1$_{InsP6}$–CTD association.

Next, we directly assessed the effect of Gle1$_{InsP6}$ on RNA release from a hydrolysis-deficient Dbp5 mutant, Dbp5(E240Q), bound to both RNA and ATP (Fig. 2f and Supplementary Table 3). Dbp5(E240Q) released RNA very slowly with a $t_{1/2}$ of 180 ± 15 s ($k_{off} = 0.0039$ s^{-1}) and Gle1$_{InsP6}$ stimulated this release ~9-fold ($t_{1/2}$ of 21 ± 6 s, $k_{off} = 0.033$ s^{-1}). Importantly, no enhancement of RNA release was observed with the Gle1(VAI→DDD) mutant (166 ± 50 s, $k_{off} = 0.042$ s^{-1}). Because mutation of Gle1 at these residues *in vivo* resulted in lethality (Supplementary Fig. 3d) and Gle1$_{InsP6}$ did not affect the ability of Dbp5 to bind or release ADP (Supplementary Table 3), we conclude that an essential function of Gle1$_{InsP6}$ is to promote RNA release from Dbp5. This may explain how Gle1$_{InsP6}$ helps to terminate mRNA export in the cytoplasm.

Given the structural similarities between Gle1$_{InsP6}$–Dbp5 and eIF4G–eIF4A, we also tested eIF4G in the same RNA release assay and found that RNA release from eIF4A is similarly stimulated in the presence of its ATPase activator (eIF4A(E172Q) alone = 55 ± 8 s, $k_{off} = 0.019 \pm 0.03$ s^{-1} compared to eIF4A(E172Q) + eIF4G = 11 ± 3 s, $k_{off} = 0.091 \pm 0.03$ s^{-1}) (Fig. 2g). This behaviour corresponds well with the ~3-fold stimulation of ATPase activity reported for eIF4G[21]. Together with the structural similarity between Gle1, eIF4G and other RNA helicase activators such as Upf2, these findings support a model in which Gle1/eIF4G-like proteins generally activate SF1 and SF2 ATPases by promoting RNA release and enzyme recycling.

Recent structural studies of the human Dbp5 orthologue, DDX19, identified an N-terminal α-helix that lowers basal Dbp5/DDX19 ATPase activity, but that is also required for RNA-stimulated ATP turnover[9,25]. Our structures lack this N-terminal helix, but structural

alignments show that Gle1$_{InsP6}$ and a C-terminal helix of Dbp5 would clash with this N-terminal domain in the RNA bound state (Supplementary Fig. 4a, b). This observation indicates that the autoinhibitory N-terminal helix must be displaced on Gle1$_{InsP6}$ association, and should have little effect on ATPase activity in the presence of Gle1$_{InsP6}$. As seen for human DDX19 (ref. 25), the yeast Δ90Dbp5 construct lacking this regulatory domain has a basal ATPase rate ~3-fold higher than full-length Dbp5. However, Δ90Dbp5 only reaches the same maximum ATPase level as full-length Dbp5 in the presence of Gle1$_{InsP6}$ (Supplementary Fig. 4c). This result indicates that release of Dbp5 autoregulation accounts for a portion of Gle1$_{InsP6}$-mediated ATPase stimulation.

Dbp5 is an ATP-dependent RNA binding protein that dissociates from RNA on each hydrolysis event (Supplementary Fig. 4d), and an interaction between full-length Dbp5 and RNA can only be detected at steady state when turnover is inhibited (Supplementary Table 3)[10]. Intriguingly, deletion of the N-terminal regulatory α-helix within Dbp5 allowed for the formation of a steady-state RNA–Dbp5 complex under conditions of hydrolysis (Supplementary Fig. 4e). These data imply that the N-terminal 90 residues of Dbp5 are involved in regulating Dbp5–RNA binding. Furthermore, the observation that Δ90Dbp5 binds RNA (Supplementary Fig. 4e), but is not stimulated for hydrolysis by RNA (Supplementary Fig. 4c), indicates that these 90 residues have a critical role in the coupling of RNA binding to ATPase activity. To test whether Gle1$_{InsP6}$ can alter steady-state RNA binding by displacing this α-helix, we added Gle1$_{InsP6}$ to RNA binding assays containing full-length Dbp5 and ATP. We observed that addition of Gle1$_{InsP6}$ allowed for steady-state binding with an apparent affinity for RNA equal to that measured for Δ90Dbp5 (Supplementary Fig. 4e and Supplementary Table 3). This observation indicates that the N-terminal extension of Dbp5 regulates RNA association, and that Gle1$_{InsP6}$ binding can alter this activity. Overall, our results are consistent with a model in which Gle1$_{InsP6}$ regulates Dbp5 by: (1) altering the regulatory function of an N-terminal helix within Dbp5; and (2) promoting substrate release through separating the Dbp5 RecA-like domains. Together, these findings explain the robust activation of Dbp5 by Gle1 and the small molecule InsP$_6$.

In addition to Gle1, Dbp5 interacts with the nucleoporin Nup159 (NUP214 in vertebrates)[3,4,8,9,26]. Nup159 and RNA bind to overlapping
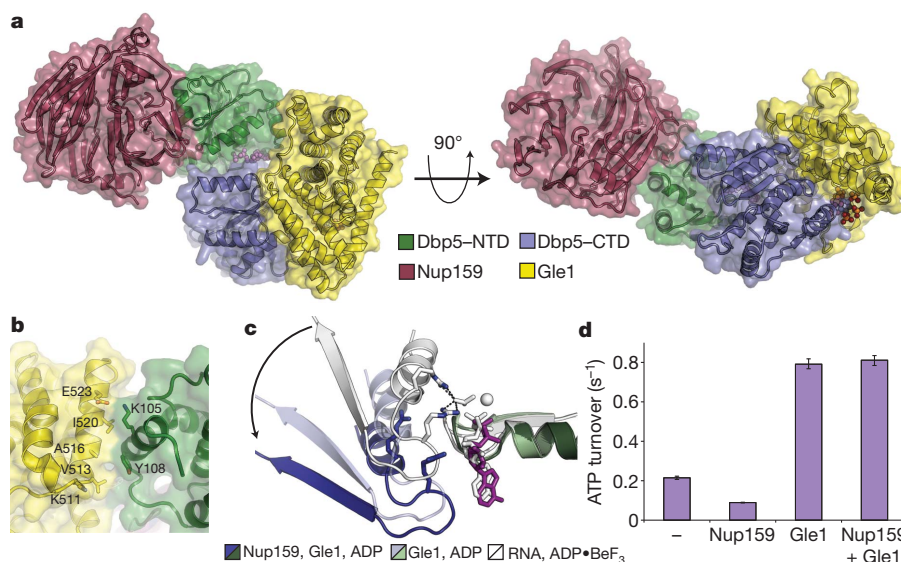
Figure 3 | The Gle1$_{InsP6}$–Δ90Dbp5–Nup159 complex. a, Two views of the Gle1$_{InsP6}$–Dbp5–Nup159 complex (see colour key). **b,** The N-terminal RecA-like domain binding interface between Dbp5–Gle1$_{InsP6}$ is altered in the presence of Nup159 (compare to Fig. 2c). Residues involved in the binding interface (sticks) are labelled. **c,** Superposition of the N-terminal RecA-domain among the three structural states of Dbp5. Arrow highlights the movement of

the CTD and catalytic arginine finger residues (sticks) among the three states (see colour key). For clarity, the Gle1–ADP intermediate is shown slightly transparent, and both the arginine finger side chains and ADP have been removed. **d,** Inhibition of the RNA-stimulated ATPase activity of Dbp5 by Nup159 is overcome in the presence of Gle1. Error bars represent s.d. ($n = 3$).
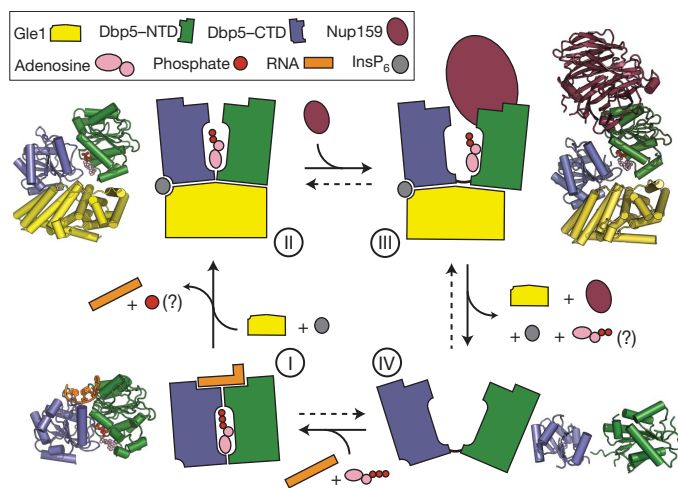
Figure 4 | Model of the Dbp5 mechanochemical cycle. In the presence of ATP, Dbp5 binds RNA causing local destabilization and remodelling of duplexed RNA or RNA–protein complexes (state I). ATP hydrolysis then allows the activator (Gle1) to bind both the C-terminal and N-terminal RecA-like domains, separating the two RecA-like domains and promoting RNA release (state II). For Dbp5 it is currently not known when inorganic phosphate (P_i) is released after hydrolysis, but this will probably occur before the formation of state II. Subsequent release of the bound RNA allows Nup159 to bind Dbp5 causing the two RecA domains to further separate (state III). The formation of this state could then facilitate ADP release, prevent rebinding of the RNA, and aid in enzyme recycling (state IV). Crystal structure model in state IV is PDB accession code 3FHO[28]. For colour coding details see key.

surfaces on Dbp5, and Nup159 inhibits RNA-dependent ATPase stimulation of Dbp5 *in vitro*[9]. To better understand how this potentially inhibitory role of Nup159 contributes to mRNA export, we solved the crystal structure of a Dbp5(L327V)–Gle1(H337R)–Nup159 complex bound to $InsP_6$ and ADP at a resolution of 2.9 Å (R_{work}/R_{free} of 22.9/26.2%) (Fig. 3a and Supplementary Table 2). Within this structure, Nup159 interacts with Dbp5 in a manner similar to that of DDX19 and NUP214 (ref. 9), occluding the RNA binding site and sterically clashing with the closed, RNA- and ATP-bound state of Dbp5 (Supplementary Fig. 5). Whereas the $Gle1_{IP6}$–CTD interface remains relatively static in all of our structures, the Gle1–NTD interface has distinct structural differences in the presence of Nup159 and now buries only 672 Å^2 of solvent-exposed surface area (compare Figs 2c and 3b). Furthermore, the two RecA-like domains of Dbp5 become further separated in the Dbp5–$Gle1_{InsP6}$–Nup159 structure, and the Dbp5 catalytic centre becomes exposed to solvent as functionally critical arginine fingers are pulled out from the nucleotide binding site (Fig. 3c). A comparison of Dbp5 among all structures reveals that interdomain structural changes proceed along a pair of related arcs, indicating that Nup159 may further assist $Gle1_{InsP6}$ in stabilizing a post-hydrolysis state of Dbp5 to allow for efficient product release at the cytoplasmic face of the NPC (Supplementary Movie 3).

To test this biochemically, we assayed Dbp5 ATPase activity and found that on addition of $Gle1_{InsP6}$, Nup159 no longer inhibits ATP turnover (Fig. 3d) indicating that $Gle1_{InsP6}$ can overcome the inhibitory effect of Nup159. This indicates that during the hydrolysis cycle in the presence of Gle1, the association of Dbp5 with Nup159 is transient, which is in agreement with the observation that the presence of adenosine nucleotide and/or RNA weakens the Dbp5–Nup159 interaction[3,9]. This provides for the possibility that Nup159 not only increases the local concentration of Dbp5 at the NPC[4,8], but also cooperates with Gle1 to promote mRNA export. Mechanistically, Gle1 and Nup159 may spatially coordinate the loading of Dbp5 onto an mRNP to target the removal of mRNA-bound export factors and contribute to the overall efficiency of mRNA export[4,8,9]. Alternatively, in a non-mutually exclusive model, the Dbp5–$Gle1_{InsP6}$–Nup159 structure may represent a post-mRNA export complex formed

after mRNP remodelling is complete and Dbp5 is removed from the mRNP. In this model, the interactions of Gle1 and Nup159 with Dbp5 would be critical for enzyme recycling and/or preventing spurious rebinding of Dbp5 to an mRNA in the terminal stages of export.

Our study reveals a surprising structural similarity between Gle1 and other helicase regulators including eIF4G and Upf2 (refs 14, 19, 21). Together with our biochemical data, these results indicate a common mode of activation for DEAD-box ATPases and RNA helicases by Gle1/eIF4G-like protein activators. Critical for the activation are a primary stable interaction between the activator and the C-terminal RecA-like domain of the ATPase, and a secondary transient interaction between the activator and the N-terminal RecA-like domain that is required to trigger nucleic-acid release (Fig. 2). For Dbp5, we propose that RNA and $Gle1_{InsP6}$/Nup159 promote the formation of closed (ATP binding and hydrolysis) and open (substrate release) conformations of the ATPase, respectively, by influencing the position of the NTD of Dbp5 relative to the CTD (Fig. 4 and Supplementary Movie 3). Complex disassembly may be further promoted by the ~50-fold ATP-dependent lowering of Dbp5 binding affinity for Nup159 observed in the presence of ATP[9], thus allowing Dbp5 to once again bind an mRNA and re-enter the ATPase cycle. For DEAD-box ATPases in general, such a toggling mechanism of the two RecA-like domains could be used to facilitate recycling of enzymes that have just melted RNA duplexes or remodelled RNA–protein complexes. Furthermore, in the presence of additional RNA-binding domains provided in *cis* or *trans*, this activity might also allow for a processive movement of this class of SF2 proteins along an RNA substrate[27].

The unique role of $InsP_6$ as a small molecule tether within the Dbp5–Gle1 complex may allow for $InsP_6$-dependent fine-tuning of interaction strength. Functionally, this could be used to modulate RNA release/turnover and consequently regulate mRNA export under various growth conditions (for example, in response to stress). Additional experiments will be needed to elucidate the functional importance of $InsP_6$ *in vivo*. However, the structural and biochemical framework presented here should guide future experiments not only to determine how Dbp5 in combination with its interaction partners confers directionality to mRNA export, but also to dissect the cellular functions of other RNA helicases and their activators.

METHODS SUMMARY

Yeast strains and plasmids are listed in Supplementary Tables 4 and 5. RNA binding and ATPase assays were performed as described in Methods. Single-stranded RNA substrates used for binding and release assays carrying a single fluorescein-conjugated UTP were made by *in vitro* transcription. RNA release assays were performed by monitoring the change in fluorescence polarization after rapid 1:1 mixing of DEAD-box protein–RNA-binding reactions with excess polyA-RNA using a stopped flow apparatus linked to a Fluoromax-3 fluorimeter. For crystallography, protein purification was performed by affinity and ion exchange chromatography followed by affinity tag cleavage and complex formation. Protein complexes were separated by gel filtration over a HiPrep Sephacryl S-200 column. Fractions containing the Dbp5–Gle1 or Nup159–Dbp5–Gle1 complex were used for crystal screening in a hanging-drop format (200 nl) using a Mosquito robotics platform at 18 °C. After 3–7 days of growth, crystals were looped and flash frozen in liquid nitrogen, and diffraction data collected on Beamline 8.3.1 at the Advanced Light Source. Detailed methods and references related to data processing, structure solution, and model building, plus all other methods used are outlined in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Cordin, O., Banroques, J., Tanner, N. K. & Linder, P. The DEAD-box protein family of RNA helicases. *Gene* **367,** 17–37 (2006).
2. Jankowsky, E. RNA helicases at work: binding and rearranging. *Trends Biochem. Sci.* **36,** 19–29 (2010).
3. Schmitt, C. *et al.* Dbp5, a DEAD-box protein required for mRNA export, is recruited to the cytoplasmic fibrils of nuclear pore complex via a conserved interaction with CAN/Nup159p. *EMBO J.* **18,** 4332–4347 (1999).

4. Hodge, C. A., Colot, H. V., Stafford, P. & Cole, C. N. Rat8p/Dbp5p is a shuttling transport factor that interacts with Rat7p/Nup159p and Gle1p and suppresses the mRNA export defect of xpo1-1 cells. *EMBO J.* **18,** 5778–5788 (1999).

5. Snay-Hodge, C. A., Colot, H. V., Goldstein, A. L. & Cole, C. N. Dbp5p/Rat8p is a yeast nuclear pore-associated DEAD-box protein essential for RNA export. *EMBO J.* **17,** 2663–2676 (1998).

6. Lund, M. K. & Guthrie, C. The DEAD-box protein Dbp5p is required to dissociate Mex67p from exported mRNPs at the nuclear rim. *Mol. Cell* **20,** 645–651 (2005).

7. Tran, E. J., Zhou, Y., Corbett, A. H. & Wente, S. R. The DEAD-box protein Dbp5 controls mRNA export by triggering specific RNA:protein remodeling events. *Mol. Cell* **28,** 850–859 (2007).

8. Weirich, C. S., Erzberger, J. P., Berger, J. M. & Weis, K. The N-terminal domain of Nup159 forms a β-propeller that functions in mRNA export by tethering the helicase Dbp5 to the nuclear pore. *Mol. Cell* **16,** 749–760 (2004).

9. von Moeller, H., Basquin, C. & Conti, E. The mRNA export protein DBP5 binds RNA and the cytoplasmic nucleoporin NUP214 in a mutually exclusive manner. *Nature Struct. Mol. Biol.* **16,** 247–254 (2009).

10. Weirich, C. S. *et al.* Activation of the DExD/H-box protein Dbp5 by the nuclear-pore protein Gle1 and its coactivator InsP$_6$ is required for mRNA export. *Nature Cell Biol.* **8,** 668–676 (2006).

11. Alcázar-Roman, A. R., Tran, E. J., Guo, S. & Wente, S. R. Inositol hexakisphosphate and Gle1 activate the DEAD-box protein Dbp5 for nuclear mRNA export. *Nature Cell Biol.* **8,** 711–716 (2006).

12. Dossani, Z. Y., Weirich, C. S., Erzberger, J. P., Berger, J. M. & Weis, K. Structure of the C-terminus of the mRNA export factor Dbp5 reveals the interaction surface for the ATPase activator Gle1. *Proc. Natl Acad. Sci. USA* **106,** 16251–16256 (2009).

13. Holm, L., Kaariainen, S., Rosenstrom, P. & Schenkel, A. Searching protein structure databases with DaliLite v.3. *Bioinformatics* **24,** 2780–2781 (2008).

14. Oberer, M., Marintchev, A. & Wagner, G. Structural basis for the enhancement of eIF4A helicase activity by eIF4G. *Genes Dev.* **19,** 2212–2223 (2005).

15. Korneeva, N. L., First, E. A., Benoit, C. A. & Rhoads, R. E. Interaction between the NH$_2$-terminal domain of eIF4A and the central domain of eIF4G modulates RNA-stimulated ATPase activity. *J. Biol. Chem.* **280,** 1872–1881 (2005).

16. Sonenberg, N. & Dever, T. E. Eukaryotic translation initiation factors and regulators. *Curr. Opin. Struct. Biol.* **13,** 56–63 (2003).

17. He, F., Brown, A. H. & Jacobson, A. Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway. *Mol. Cell. Biol.* **17,** 1580–1594 (1997).

18. Chamieh, H., Ballut, L., Bonneau, F. & Le Hir, H. NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nature Struct. Mol. Biol.* **15,** 85–93 (2008).

19. Kadlec, J., Guilligay, D., Ravelli, R. B. & Cusack, S. Crystal structure of the UPF2-interacting domain of nonsense-mediated mRNA decay factor UPF1. *RNA* **12,** 1817–1824 (2006).

20. Alcázar-Román, A. R., Bolger, T. A. & Wente, S. R. Control of mRNA export and translation termination by inositol hexakisphosphate requires specific interaction with Gle1. *J. Biol. Chem.* **285,** 16683–16692 (2010).

21. Schutz, P. *et al.* Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein–protein interactions. *Proc. Natl Acad. Sci. USA* **105,** 9564–9569 (2008).

22. Marintchev, A. *et al.* Topology and regulation of the human eIF4A/4G/4H helicase complex in translation initiation. *Cell* **136,** 447–460 (2009).

23. Henn, A., Cao, W., Hackney, D. D. & De La Cruz, E. M. The ATPase cycle mechanism of the DEAD-box rRNA helicase, DbpA. *J. Mol. Biol.* **377,** 193–205 (2008).

24. Henn, A. *et al.* Pathway of ATP utilization and duplex rRNA unwinding by the DEAD-box helicase, DbpA. *Proc. Natl Acad. Sci. USA* **107,** 4046–4050 (2010).

25. Collins, R. *et al.* The DEXD/H-box RNA helicase DDX19 is regulated by an α-helical switch. *J. Biol. Chem.* **284,** 10296–10300 (2009).

26. Napetschnig, J. *et al.* Structural and functional analysis of the interaction between the nucleoporin Nup214 and the DEAD-box helicase Ddx19. *Proc. Natl Acad. Sci. USA* **106,** 3089–3094 (2009).

27. Büttner, K., Nehring, S. & Hopfner, K. P. Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nature Struct. Mol. Biol.* **14,** 647–652 (2007).

28. Fan, J. S. *et al.* Solution and crystal structures of mRNA exporter Dbp5p and its interaction with nucleotides. *J. Mol. Biol.* **388,** 1–10 (2009).

Author Contributions B.M., N.D.T., J.M.B. and K.W. designed the experiments. Protein purifications for crystallization and biochemical assays were performed by B.M. Crystal screening, optimization and data collection were performed by B.M. and N.D.T. Data processing, structure solution and model building was performed by N.D.T. *In vivo* tests and *in vitro* RNA binding, ATPase assays and other biochemical assays were performed by B.M. with assistance from M.A.S. and K.J.H. B.M., N.D.T., J.M.B. and K.W. both analysed and interpreted the data and wrote the manuscript.

Author Information Coordinates and structure factors for Dbp5–CTD(L327V)/InsP$_6$/Gle1(H337R), Dbp5–CTD(L327V)/InsP$_6$/Gle1(WT), Δ90Dbp5(L327V)/RNA/ADP•BeF$_3$, Δ90Dbp5/RNA/ADP•BeF$_3$, Δ90Dbp5(L327V)/InsP$_6$/Gle1(H337R)/ADP and Δ90Dbp5(L327V)/InsP$_6$/Gle1(H337R)/ADP/Nup159 have been deposited in the protein data bank under accession numbers 3PEU, 3PEV, 3PEW, 3PEY, 3PEX and 3PEZ, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.W. (kweis@berkeley.edu) or J.M.B. (jmberger@berkeley.edu).

METHODS

Protein purification for crystallography. Dbp5, Gle1 (amino acids (a.a.) 244–538), Nup159 (a.a. 2–387), eIF4A and eIF4G (a.a. 572–952) from *Saccharomyces cerevisiae* were expressed as previously described[8,10,21] in BL21-CodonPlus (DE3)-RIPL cells (Strategene) using a combination of nickel affinity, ion exchange and size exclusion chromatography. For the production of selenomethionine-labelled protein, Dbp5 and Gle1 proteins were purified as above, except that cells were grown in minimal media supplemented with selenomethionine and β-ME and dithiothreitol (DTT) were replaced with tris(2-carboxyethyl)phosphine (TCEP) throughout the purification. TEV cleavage was done at a 1:50 ratio (w/w) in 30 mM HEPES (pH 7.5), 400 mM NaCl, 1 mM DTT, 0.25 mM InsP$_6$ and 10% glycerol at 22 °C for 4 h[29]. After removal of uncleaved protein, by orthogonal purification over a mixture of HIS-Select HF Nickel Affinity Gel (Sigma) and amylose resin (NEB), proteins were mixed in an equimolar ratio, diluted to 150 mM NaCl and concentrated to ~15 mg ml^{-1} at 4 °C (Amicon Ultra; Millipore). The protein mixture was then further separated by gel filtration over a HiPrep Sephacryl S-200 column (GE Healthcare) at 0.4 ml min^{-1} in 10 mM HEPES (pH 7.5), 100 mM NaCl, 1 mM DTT, 0.50 mM InsP$_6$ and 5% glycerol at 4 °C. Fractions containing the desired protein complex, in an equimolar ratio (as judged by SDS–PAGE), were then pooled and concentrated to ~25 mg ml^{-1} at 4 °C.

Protein crystallization. Protein solutions were stored at 4 °C and used for crystallization within 1 week, during which time there was no detectable degradation of the proteins. All crystallization was performed in a hanging drop format using a Mosquito robotics platform (TTP LabTech) at 18 °C.

The Dbp5–CTD(L327V)/InsP$_6$/Gle1(H337R) crystals were grown by mixing 100 nl of selenomethionine-labelled protein solution at 14 mg ml^{-1} in gel filtration buffer with 100 nl of a solution containing 30% PEG 3350, 100 mM HEPES pH 8.0, 50 mM NaOAc and 200 mM LiSO$_4$, and incubating for 5 days. Crystals were cryoprotected by dilution into a solution containing 24% PEG 3350, 80 mM HEPES 8.0, 160 mM LiSO$_4$, 40 mM NaOAc, 20% glycerol, 0.25 mM InsP$_6$, 1 mM DTT and 100 mM NaCl, and flash frozen in liquid nitrogen.

The Dbp5–CTD(L327V)/InsP$_6$/Gle1(WT) crystals were grown using the same methods as for Dbp5–CTD(L327V)/InsP$_6$/Gle1(H337R), but proteins were subjected to gel filtration individually before forming 1:1 complexes by direct mixing. Crystallization was performed by mixing 20 mg ml^{-1} of selenomethionine-labelled protein solution in gel filtration buffer (with NaCl increased to 150 mM) with a solution containing 30% PEG 3350, 100 mM HEPES 7.8 and 200 mM LiSO$_4$ and incubating for 5 days. Crystals were cryoprotected by diluting into a solution containing 24% PEG 3350, 80 mM HEPES 7.8, 160 mM LiSO$_4$, 10% glycerol, 0.4 mM InsP$_6$, 1 mM DTT and 120 mM NaCl, and flash frozen in liquid nitrogen.

The Δ90Dbp5(L327V)/RNA/ADP•BeF$_3$ complex was formed by mixing a solution containing unlabelled Δ90Dbp5(L327V) in gel filtration buffer (with the addition of 5 mM MgCl$_2$) with a 1.2:1 molar ratio of rU$_{10}$ RNA (IDT), incubating for 15 min, and then mixing with ADP•BeF$_3$ (prepared in a 1:3:15 ratio of ADP:Be:F) to a final concentration of 1 mM, followed by another 15 min of incubation. Crystals of the Δ90Dbp5(L327V)/RNA/ADP•BeF$_3$ complex were grown by mixing 100 nl of the complex at a protein concentration of 15 mg ml^{-1} with 100 nl of a solution containing 200 mM MgNO$_3$ and 20% PEG 3350 (The PEGs Suite, condition #64; Qiagen) and incubating for 7 days. Crystals were cryoprotected by diluting in a solution containing 18% PEG 3350, 9 mM HEPES pH 7.5, 180 mM MgNO$_3$, 4.5% glycerol, 1 mM DTT, 90 mM NaCl, 0.5 mM InsP$_6$, 4.5 mM MgCl$_2$ and 1 mM ADP•BeF$_x$ (prepared in a 1:3:15 ratio of ADP:Be:F), and flash frozen in liquid nitrogen. The Δ90Dbp5/RNA/ADP•BeF$_3$ crystals were grown in the same manner using wild type Δ90Dbp5.

The Δ90Dbp5(L327V)/InsP$_6$/Gle1(H337R)/ADP complex was formed by mixing the unlabelled protein solution with ADP to a final concentration of 0.5 mM and incubated for 15 min. Crystals were grown by mixing 200 nl of a solution containing 30% PEG 300 and 100 mM MES pH 6.5 with 50 nl of a 10% MPD additive solution, adding 250 nl of protein–ADP complex at 15 mg ml^{-1} protein in gel filtration buffer (with the addition of 10 mM MgCl$_2$), and incubating for 9 days. Crystals were cryoprotected by diluting in a solution containing 27% PEG 300, 90 mM MES pH 6.5, 9 mM HEPES pH 7.5, 5.4% MPD, 90 mM NaCl, 9 mM MgCl$_2$, 0.9 mM DTT, 0.9 mM ADP, 0.45 mM InsP$_6$ and 4.5% glycerol and flash frozen in liquid nitrogen.

The Nup159/Δ90 Dbp5(L327V)/InsP$_6$/Gle1(H337R)/ADP complex was formed by mixing the unlabelled protein solution with ADP to a final concentration of 1.0 mM, and incubating for 15 min. Crystals were grown by mixing 200 nl of a solution containing 20% PEG 33350 and 200 mM KOAc with 50 nl of a 100 mM sarcosine additive solution, adding 250 nl of protein–ADP complex at 15 mg ml^{-1} protein in gel filtration buffer (with the addition of 10 mM MgCl$_2$), and incubating for 5 days. Crystals were cryoprotected by diluting in a solution containing 27% PEG 3350, 180 mM KOAc, 90 mM NaCl, 9 mM HEPES pH 7.5, 9 mM sarcosine, 9 mM MgCl$_2$, 0.9 mM DTT, 0.9 mM ADP, 0.45 mM InsP$_6$ and 4.5% glycerol and flash frozen in liquid nitrogen.

Data collection, structure solution and refinement. All data were collected at beamline 8.3.1 at the Advanced Light Source at 100 Kelvin[30], and processed using either HKL2000 or ELVES[31,32] (see also later). Structures were solved with the AutoSol or AutoMR components of PHENIX[33]. Iterative rounds of manual model building in COOT[34], coupled with refinement using PHENIX[33] and stereochemical validation using MOLPROBITY[35], resulted in final models displaying excellent geometry and R_{work}/R_{free} values for their respective resolutions, as well as no Ramachandran outliers. Rigid body refinement was conducted on Gle1 and Dbp5 RecA-like domains for all structures solved by molecular replacement in order to reveal any rigid-body differences among the structures. Simulated annealing omit maps were calculated to confirm the presence of all ligands (Supplementary Fig. 6). For all structures solved by molecular replacement, simulated annealing composite omit maps and prime and switch maps, both made using the AutoBuild component of PHENIX[33], were calculated to help remove model bias.

MAD data were collected at wavelengths of 0.9796, 0.9798 and 1.020 Å for the Dbp5–CTD(L327V)/InsP$_6$/Gle1(H337R) crystals in inverse beam mode and processed in HKL2000[31] (Supplementary Table 1). The structure was solved using PHENIX AutoSol, which found 11 selenium sites and gave figures of merit of 0.5 and 0.7 for phasing and density modification, respectively. An initial model was built using PHENIX AutoBuild[33]. For the final model, atomic coordinates, individual atomic displacement parameters (B-factors), translation/libration/screw (TLS) parameters, and f' and f'' parameters for the Se atoms were refined against the unmerged Friedel pairs (F+ and F−) of the remote data set as implemented in PHENIX. Ramachandran statistics are 98.3% preferred, 1.7% allowed and 0% outliers as reported by Molprobity[35].

A single low-energy remote data set was collected at a wavelength of 1.116 Å for the Dbp5–CTD(L327V)/InsP$_6$/Gle1(WT) crystals and processed in HKL2000 with merging of Friedel pairs. The structure was solved using PHENIX AutoMR by searching for a single copy of the fully refined Dbp5–CTD(L327V)/InsP$_6$/Gle1(H337R) structure[33]. For the final model, atomic coordinates, individual atomic displacement parameters (B-factors), and TLS parameters were refined. The refinement test set was copied from the isomorphous Dbp5–CTD(L327V)/InsP$_6$/Gle1(H337R) data and extended to higher resolution in order to avoid biasing the R_{free}. Ramachandran statistics are 98.0% preferred, 2.0% allowed and 0% outliers as reported by Molprobity[35].

Native data were collected on the Δ90Dbp5(L327V)/RNA/ADP•BeF$_3$ crystals at a wavelength of 1.116 Å and processed in HKL2000. The structure was solved using PHENIX AutoMR by searching for a single copy of a *S. cerevisiae* Dbp5 homology model (SWISS-MODEL)[33,36], created using the human Ddx19/RNA/AMPPNP structure as a template (PDB accession code 3G0H)[25]. For the final model, atomic coordinates, B-factors, TLS parameters and occupancies for residues with alternate conformations were refined. Custom bond restraints were used for octahedral Mg^{2+}-H$_2$O clusters and the ADP•BeF$_3$ molecule. Ramachandran statistics are 99.2% preferred, 0.8% allowed and 0% outliers as reported by Molprobity[35].

Native data were collected on the Δ90Dbp5(WT)/RNA/ADP•BeF$_3$ crystals at a wavelength of 1.116 Å and processed in HKL2000. The structure was solved using PHENIX AutoMR by searching for a single copy of the fully refined Δ90Dbp5(L327V)/RNA/ADP•BeF$_3$ structure and refined in the same manner[33]. Ramachandran statistics are 98.5% preferred, 1.5% allowed and 0% outliers as reported by Molprobity[35].

Native data were collected on the Δ90Dbp5(L327V)/InsP$_6$/Gle1(H337R)/ADP crystals at a wavelength of 1.116 Å and processed in Elves. The structure was solved using PHENIX AutoMR by searching for a single copy of the fully refined Gle1(H337R)/InsP$_6$/Dbp5–CTD(L327V) structure and a single copy of the Dbp5–NTD (residues 101–293) of the fully refined Δ90Dbp5(L327V)/RNA/ADP•BeF$_3$ structure[33]. B-factor sharpening was used throughout model building. Initial $F_o − F_c$ difference density showed that two additional InsP$_6$ molecules were present in the structure. Although the InsP$_6$ molecules have high relative B-factors, they mediate crystal-packing contacts and are coordinated by a large number of basic and polar residues. For the final structure, atomic coordinates and TLS parameters were refined. Secondary structure restraints as implemented in PHENIX were used throughout refinement to maintain proper backbone geometry. Ramachandran statistics are 97.6% preferred, 2.4% allowed and 0% outliers as reported by Molprobity[35].

Native data were collected on the Nup159/Δ90Dbp5(L327V)/InsP$_6$/Gle1(H337R)/ADP crystals at a wavelength of 1.116 Å and processed in HKL2000. The structure was solved using PHENIX AutoMR by searching for a single copy each of Gle1, the Dbp5–NTD, the Dbp5–CTD and Nup159 (PDB accession code 1XIP)[8,33]. For the

final model, atomic coordinates and TLS parameters were refined. Ramachandran statistics are 97.3% preferred, 2.7% allowed and 0% outliers as reported by Molprobity[35].

Structural analysis. All structural analysis and rendering for figures was performed with PyMOL (Schrödinger, LLC). Solvent accessible electrostatics in Fig. 1 was calculated using ABPS[37]. Energy minimized linear interpolations shown in Supplementary Movies 1–3 were made using a CNS script written by the Yale Morph Server[38,39]

Complex formation and gel filtration. To assay for Dbp5-Gle1 complex formation, 100 μl of a binding reaction was applied to a Superdex 200 gel filtration column (GE Healthcare) in the presence of 30 mM HEPES (pH 7.5), 150 mM NaCl, 1 mM DTT, 0.25 mM InsP$_6$ and 5% glycerol at 0.4 ml min^{-1}. Binding reactions contained Dbp5 (20 μM) and Gle1 (40 μM) mixed with column buffer.

ATPase assays. ATPase assays using Dbp5 were performed as previously described[10]. When present, Nup159 was added to a final concentration of 1 μM.

RNA binding and release assay. The RNA substrate used for the binding and release assays was a 29-nucleotide single-stranded (ss)RNA of the sequence 5′-GGGUAAAAAAAAAAAAAAAAAAAAAAAAAAA-3′ for Dbp5 or a 25-nucleotide ssRNA of the sequence 5′-GGGUACAACGGAAGACAGCAGAGAA for eIF4A carrying a fluorescein-conjugated UTP made by *in vitro* transcription (MEGAshortscript Kit, Ambion). RNA binding assays were performed as previously described[10,40]. To perform RNA release assays, Dbp5(E240Q) (1.0 μM) or eIF4A(E172Q) (4.0 μM), FITC-RNA (40 nM), and ATP (2 mM) ± Gle1 (2.0 μM) or eIF4G (8.0 μM) were complexed for 5 min in a buffer containing 50 mM HEPES (pH 7.5), 140 mM KCl, 5 mM MgCl$_2$, 1 mM DTT, 20% glycerol, 0.1 mg ml^{-1} BSA and 10 μM InsP$_6$ for Dbp5 or 20 mM MES (pH 6.0), 20 mM KOAc, 5 mM MgCl$_2$, 1 mM DTT, 20% glycerol, and 0.1 mg ml^{-1} BSA for eIF4A. After rapid 1:1 mixing of DEAD-box-RNA binding reactions with polyA-RNA (1.0 mg ml^{-1}) using a stopped flow apparatus (RX2000; Applied Photophysics) linked to a Fluoromax-3 fluorimeter (Horiba Jobin Yvon), fluorescence intensities were measured every 0.05 to 0.2 s depending on the reaction. To measure fluorescent polarization, samples were excited with vertically polarized light (492 nM), and emission (521 nM) of vertically (I_{VV}) and horizontally (I_{VH}) polarized light was measured independently in two consecutive runs. Anisotropy (r) was then calculated using the equation: $r = (I_{VV} - (G \times I_{VH}))/(I_{VV} + (2 \times G \times I_{VH}))$, where $G = I_{HV}/I_{HH}$. Reported $t_{1/2}$ values are an average of three independent experiments ± standard deviation. Release curves were fit to a first order exponential decay curve using KaleidaGraph (Synergy Software).

ADP binding and release assays. 2′-(or-3′)-O-(N-methylanthraniloyl)adenosine 5′-diphosphate (MANT-ADP) (Invitrogen) was used for the binding and release assays with Dbp5(E240Q). MANT-ADP binding assays were performed in a manner similar to the RNA binding assay[10]. Briefly, Dbp5(E240Q) (0–2 μM) and MANT-ADP (100 nM) were mixed with buffer (50 mM HEPES pH 7.5, 140 mM KCl, 5 mM MgCl$_2$, 1 mM DTT, 20% glycerol, 0.1 mg ml^{-1} BSA and 10 μM InsP$_6$) and RNA (1 mg ml^{-1}) or Gle1 (2 μM) as indicated. After mixing, anisotropy was measured using a Fluoromax-3 fluorimeter (Horiba Jobin Yvon). To perform the release assay, Dbp5(E240Q) (1 μM) was complexed with MANT-ADP (100 nM) for 5 min ± Gle1 (2 μM) and/or RNA (1 mg ml^{-1}) in a buffer containing 50 mM HEPES pH 7.5, 140 mM KCl, 5 mM MgCl$_2$, 1 mM DTT, 20% glycerol, 0.1 mg ml^{-1} BSA and 10 μM InsP$_6$. After rapid 1:1 mixing of Dbp5-MANT-ADP reactions with excess ADP (2 mM) using a stopped flow apparatus (RX2000; Applied Photophysics) linked to a Fluoromax-3 fluorimeter (Horiba Jobin Yvon), fluorescence intensities were measured every 0.01 s using 370 nM (excitation) and 445 nM (emission) light. Fluorescent polarization and $t_{1/2}$ values were calculated as described earlier for the RNA release assays.

29. Kapust, R. B. & Waugh, D. S. Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expr. Purif.* **19,** 312–318 (2000).
30. MacDowell, A. A. *et al.* Suite of three protein crystallography beamlines with single superconducting bend magnet as the source. *J. Synchrotron Radiat.* **11,** 447–455 (2004).
31. Otwinowski, Z. & Minor, W. in *Methods in Enzymology* (eds Carter, W. C. & Sweet, R. M.) 307–326 (Academic, 1997).
32. Holton, J. & Alber, T. Automated protein crystal structure determination using ELVES. *Proc. Natl Acad. Sci. USA* **101,** 1537–1542 (2004).
33. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
34. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
35. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35,** W375–W383 (2007).
36. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22,** 195–201 (2006).
37. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98,** 10037–10041 (2001).
38. Flores, S. *et al.* The Database of Macromolecular Motions: new features added at the decade mark. *Nucleic Acids Res.* **34,** D296–D301 (2006).
39. Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
40. Lorsch, J. R. & Herschlag, D. The DEAD box protein eIF4A. 1. A minimal kinetic and thermodynamic framework reveals coupled binding of RNA and nucleotide. *Biochemistry* **37,** 2180–2193 (1998).

LETTER

Structure and mechanism of the uracil transporter UraA

Feiran Lu[1]*, Shuo Li[1]*, Yang Jiang[1], Jing Jiang[1], He Fan[1], Guifeng Lu[1], Dong Deng[1], Shangyu Dang[1], Xu Zhang[1], Jiawei Wang[1] & Nieng Yan[1]

The nucleobase/ascorbate transporter (NAT) proteins, also known as nucleobase/cation symporter 2 (NCS2) proteins, are responsible for the uptake of nucleobases in all kingdoms of life and for the transport of vitamin C in mammals[1,2]. Despite functional characterization of the NAT family members in bacteria, fungi and mammals, detailed structural information remains unavailable. Here we report the crystal structure of a representative NAT protein, the *Escherichia coli* uracil/H^+ symporter UraA, in complex with uracil at a resolution of 2.8 Å. UraA has a novel structural fold, with 14 transmembrane segments (TMs) divided into two inverted repeats. A pair of antiparallel β-strands is located between TM3 and TM10 and has an important role in structural organization and substrate recognition. The structure is spatially arranged into a core domain and a gate domain. Uracil, located at the interface between the two domains, is coordinated mainly by residues from the core domain. Structural analysis suggests that alternating access of the substrate may be achieved through conformational changes of the gate domain.

Representative NAT/NCS2 family proteins include the vitamin C transporters SVCT1 and SVCT2 in mammals[3–5]; the uracil transporter SNBT1 in rat[6]; the uric acid/xanthine transporter UapA in *Aspergillus nidulans*[7]; the UapA homologue, YgfO (also known as XanQ), in *E. coli*[8]; and the uracil and 5-flurouracil transporter UraA[9] (Supplementary Fig. 1). NAT proteins contain a signature motif, (A, G, S)(Q, E, P)N-X-G-X$_4$-T(R, K, G), where X denotes a nonspecific amino acid, that may be involved in substrate recognition and transport[10–12]. Whereas the mammalian NAT proteins are sodium symporters[6], their bacterial homologues are proton symporters[2]. Uptake of nucleobases can also be mediated by members of the nucleobase/cation symporter 1 (NCS1) family[1,13], which share little sequence similarity with NAT/NCS2 proteins. The NCS1 transporter Mhp1 is structurally similar to LeuT[14–16]. It remains to be seen whether the NAT/NCS2 transporters also conform to the LeuT fold.

Uracil binds to recombinant UraA protein with a dissociation constant of approximately $0.41 \pm 0.07\ \mu M$ (s.d. of three independent experiments) as measured by scintillation proximity assay[17] (Supplementary Fig. 2a). Uracil proved to be essential for generation of usable UraA crystals grown in the presence of *n*-nonyl-β-D-glucopyranoside (β-NG). We determined the structure using mercury-based single isomorphous replacement (Supplementary Table 1). The electron density was of excellent quality (Supplementary Fig. 3a) except for that of residues 179–195, which were modelled as poly-Ala. During structure refinement, a prominent, disc-shaped electron density reminiscent of uracil appeared in the centre of UraA after most amino acids were modelled (Supplementary Fig. 3b). Because crystal growth strictly depended on the inclusion of 1 mM uracil in the protein purification and crystallization solution, we modelled uracil into the density and refined the final atomic model at a resolution of 2.8 Å (Supplementary Table 1 and Supplementary Fig. 3c).

UraA contains 14 transmembrane segments, with both the amino and the carboxy termini located on the cytoplasmic side (Fig. 1a). The two C-terminal α-helices, α13 and α14, are only halfway into the membrane. There is no detectable structural similarity between UraA and Mhp1[14]. Searching the Protein Data Bank using the Dali server[18] suggested that UraA has a novel fold (Supplementary Fig. 4a). Notably, two short, antiparallel β-strands on TM3 and TM10 are located at the centre of the structure. Each β-strand is preceded by an extended, unwound fragment and followed by a short α-helix, α3 or α10 (Fig. 1b). Discontinuous helices have been observed in a number of transport proteins[19], but it is unusual for an unwound region to constitute half of the transmembrane segment. These structural features distinguish UraA from other known structures of integral membrane proteins. The 11-residue NAT signature motif, which was predicted to be a loop in the cytoplasm[10,12], constitutes membrane-embedded α10 in UraA (Supplementary Fig. 1).

The 14 transmembrane segments of UraA are arranged into two structural repeats, TM1–TM7 and TM8–TM14, which are related to each other by a rotation of approximately 180° around an axis parallel to the membrane bilayer. These two repeats can be superimposed with a root mean squared deviation of 2.9 Å over 135 aligned Cα atoms (Supplementary Fig. 4b). The presence of two inverted repeats is a shared feature for a number of transporters[16] and channels[20,21].

The 14 transmembrane segments are spatially organized into a core domain and a gate domain. The core domain comprises TM1–TM4 and TM8–TM11, and the gate domain contains the other six segments (Fig. 2a, left). The interface between the two domains is populated mainly with hydrophobic residues from TM1, TM3 and TM8 in the core domain and TM5 and TM12 in the gate domain (Fig. 2a, right). Notably, the interdomain interactions are further stabilized by the head group of a bound β-NG molecule (Supplementary Fig. 5). We added β-NG in the last step of protein purification, whereas uracil was included throughout purification and crystallization. Uracil binds UraA in the absence of β-NG (Supplementary Fig. 2a), whereas addition of β-NG to substrate-free UraA led to severe protein precipitation that prevented biochemical analysis. Furthermore, D-glucose, the head group of β-NG, had no impact on uracil binding or transport when measured by scintillation proximity assay or cell-uptake assay. Thus, we concluded that uracil-bound UraA may provide a suitable conformation for β-NG to bind to and that β-NG binding is unlikely to have an impact on uracil recognition.

In contrast to the hydrophobic interface between the two domains, there are a large number of buried hydrogen bonds within the core domain, with the β-strands of TM3 and TM10 at the centre (Fig. 2b). Apart from conventional hydrogen bonds in antiparallel β-strands, the hydroxyl group of Tyr 288 in TM10 makes two hydrogen bonds with the carbonyl oxygen and amide nitrogen of Ser 71 in TM3. There are 12 additional hydrogen bonds between TM3/TM10 and other transmembrane segments (Fig. 2b). These extensive interactions may facilitate the conformation of the unwound regions in TM3 and TM10 as well as restraining movement of the transmembrane segments in the core domain.

[1]State Key Laboratory of Bio-membrane and Membrane Biotechnology, Center for Structural Biology, School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China.
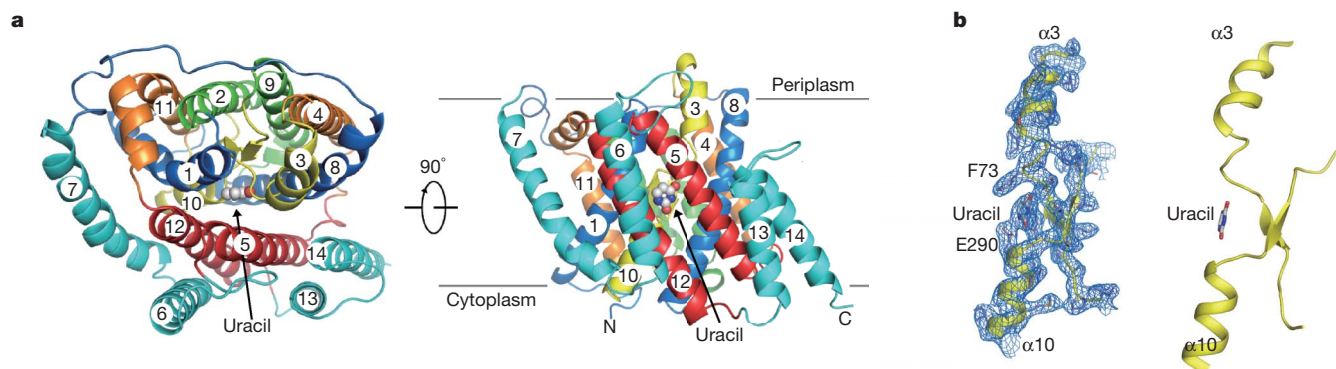*These authors contributed equally to this work.

Figure 1 | The structure of UraA reveals a novel fold. a, Overall structure of UraA. Two perpendicular views, one from the periplasm and one from the side, are shown. The bound uracil is indicated. **b**, A pair of short, antiparallel β-strands is located in the middle of TM3 and TM10. The $2F_o - F_c$ electron density map, shown in blue mesh, is contoured at 1σ on the left. All structural figures, including the calculation of surface electrostatic potential, were prepared with PYMOL[30].

The antiparallel β-strands and the connecting loops of TM3 and TM10 provide a shelter for uracil between the core and gate domains. The pyrimidine ring of uracil, which is roughly parallel to the β-strands, is surrounded by negative electrostatic potentials (Fig. 3a). Recognition of uracil is almost exclusively mediated by the core domain, involving residues from TM1, TM3, TM8, TM10 and TM12 (Fig. 3b and Supplementary Fig. 4c). Two Glu residues, Glu 241 and Glu 290, anchor uracil by each making two hydrogen bonds with it (Fig. 3b). Replacement of either Glu residue by Ala completely abrogated uracil binding (Supplementary Fig. 2b). In addition, the two oxygen atoms of uracil

form hydrogen bonds with the amide nitrogen atoms of Phe 73 and Gly 289. Notably, Gly 289 and Glu 290 are within the NAT signature motif. A hydroxyl group in β-NG forms hydrogen bonds with both the imidazole nitrogen of His 245 and the carbonyl oxygen of uracil (Supplementary Fig. 5b), suggesting that a water molecule may occupy the position of the β-NG hydroxyl group and make similar interactions. In this case, His 245 may contribute to uracil binding with a water-mediated hydrogen bond (Supplementary Fig. 5c). Consistent with this notion, substitution of Ala for His 245 abolished uracil binding (Supplementary Fig. 2b).
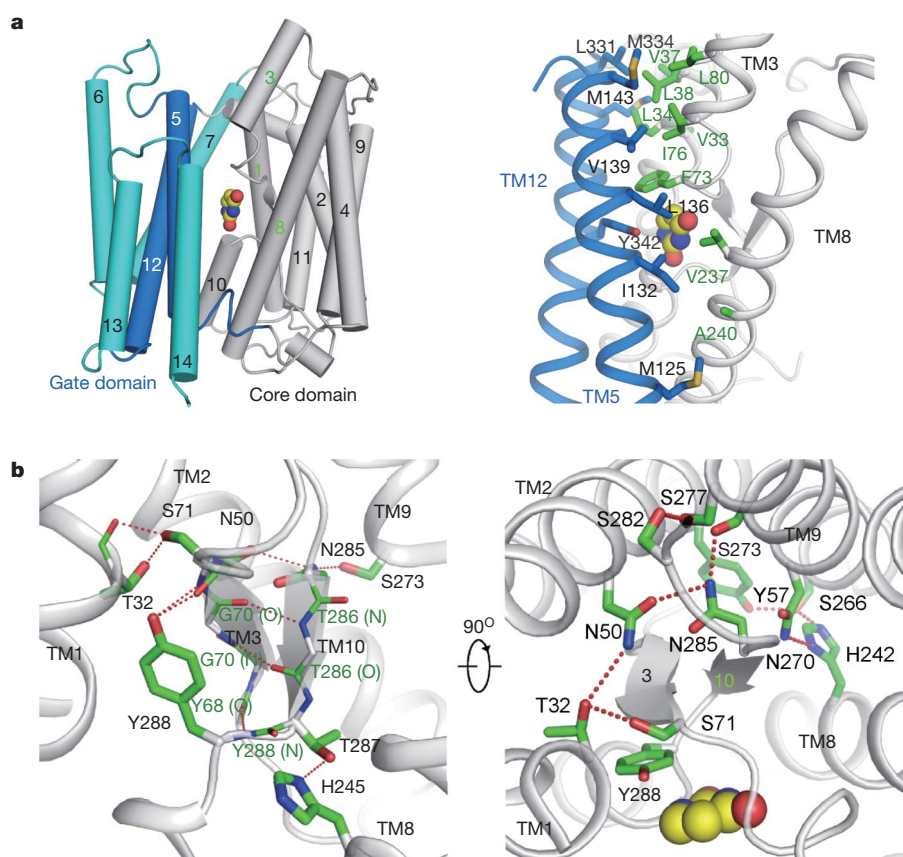

Figure 2 | Domain organization of UraA. a, Left: UraA is spatially organized into core (grey) and the gate (cyan and blue) domains that associate with each other through hydrophobic interactions. Right: the residues that mediate the interdomain interactions are shown as green (core) and blue (gate) sticks in the right panel. Uracil is shown as yellow spheres. **b**, Antiparallel β-strands in TM3 and TM10 provide the organizing centre for the core domain. There are a large number of hydrogen bonds between transmembrane segments within the core domain. Two perpendicular views are shown. Hydrogen bonds are represented by red dashed lines.
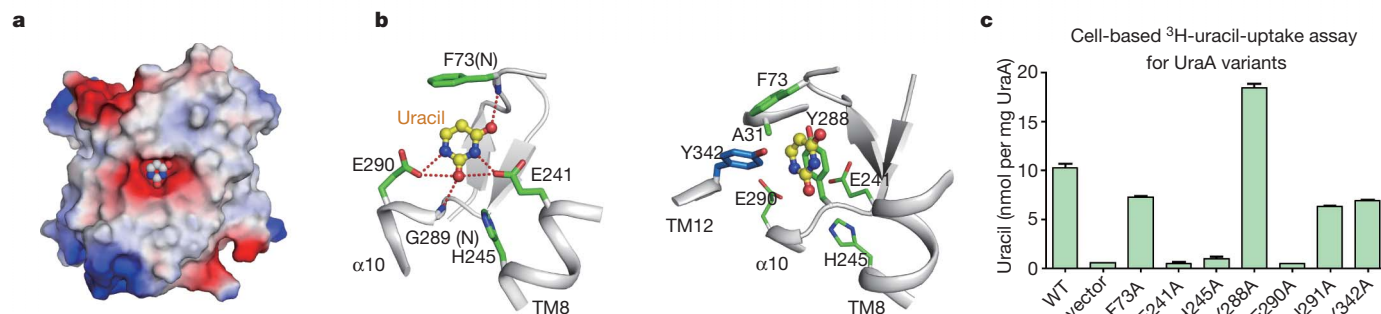
Figure 3 | Uracil coordination by UraA. a, Uracil is located in a concave pocket surrounded by negative electrostatic potential. The core domain of UraA is shown in surface electrostatic potential. **b**, Uracil is coordinated by both polar (left) and van der Waals (right) contacts. Uracil is shown in yellow, in ball-and-stick form. **c**, Cell-based ^{3}H-uracil-uptake assay identified the key residues in uracil transport. The membrane expression levels of UraA variants were monitored by western blot using an antibody against the His tag. The amount of each protein was estimated by comparing the intensity against a serial dilution of UraA with known concentration on the same western blot. The reactions lasted for 30 s. Details of the experiments are described in Supplementary Fig. 6 and Methods. WT, wild type. Error bars, s.d. of three independent experiments.

In addition to hydrogen bonds, uracil is coordinated by van der Waals interactions involving Ala 31, Phe 73, Tyr 288 and Tyr 342. Phe 73 blocks access to uracil from the periplasm, whereas the phenyl ring of Tyr 288 is roughly parallel to the pyrimidine ring of uracil. Tyr 342 is the only residue from the gate domain that contributes to uracil binding (Fig. 3b). The presence of aromatic residues surrounding the substrate is commonly observed in membrane transporters, such as the sodium-coupled sugar transporter vSGLT[22], the glycine betaine transporter BetP[23] and the arginine/agmatine antiporter AdiC[24,25]. The bulky residues effectively insulate the substrate from the outside environment.

To corroborate structural observations, we generated a number of UraA variants, each containing a mis-sense mutation, and examined their ability to transport uracil into *E. coli*. Whereas the wild-type UraA can mediate the uptake of ^{3}H-uracil with a Michaelis constant, K_m, of approximately 0.5 μM (Supplementary Fig. 6), replacement of Glu 241, His 245 or Glu 290 by Ala invariably abrogated uracil uptake (Fig. 3c). This finding is consistent with loss of uracil binding by these three mutants (Supplementary Fig. 2). By contrast, the UraA variants Phe 73 Ala and Tyr 342 Ala retained the bulk of transport activities (Fig. 3c).

Mutation of the invariant Asn residue in the NAT motif was shown to abrogate the transport function for UapA and YgfO[10–12]. In UraA, however, only Gly 289 and Glu 290 of the NAT motif are directly involved in substrate binding. Asn 291 is located away from the substrate-binding site (Supplementary Fig. 7). The UraA variant Asn 291 Ala retained the majority of the wild-type uracil-uptake activity (Fig. 3c). Because UapA

and YgfO are xanthine/uric acid permease, the invariant Asn residue in the NAT motif may have a more important role in these purine transporters.

UraA is a proton-coupled symporter. Translocation of proton relies on the residues that can be protonated and deprotonated. Remarkably, Glu 241, His 245 and Glu 290, all of which can undergo cycles of protonation and deprotonation, are clustered at the interface between the core and gate domains (Supplementary Fig. 8) and are essential for uracil binding (Fig. 3b and Supplementary Fig. 2b). This arrangement suggests that these residues may have a key role in proton translocation. Curiously, however, only Glu 241 is invariant among all NAT members. His 245 is replaced by Asp, Thr or Val in other NAT proteins, whereas Glu 290 is substituted by Gln or Pro (Supplementary Fig. 1). A closer examination revealed a pattern of function-based conservation. Glu 290 is conserved in the three known pyrimidine permeases RutG, PyrP and SNBT1. His 245 is conserved in RutG and PyrP, both of which are pyrimidine/proton symporters. By contrast, His 245 is replaced by a conserved Asp residue in the sodium symporters SNBT1, SVCT1 and SVCT2 (Supplementary Fig. 1). This analysis suggests conserved mechanisms of transport, with His 245 having a key role in proton symport and Asp having a similar role in sodium symport.

An analysis of solvent accessibility using the program HOLE[26] unambiguously shows that uracil is occluded from the periplasm but readily accessible from the cytoplasmic side (Fig. 4a). Thus, our current structure of uracil-bound UraA adopts an inward-open conformation. To load uracil, UraA must also be able to adopt an outward-open conformation. This is probably accomplished by a rigid-body rotation
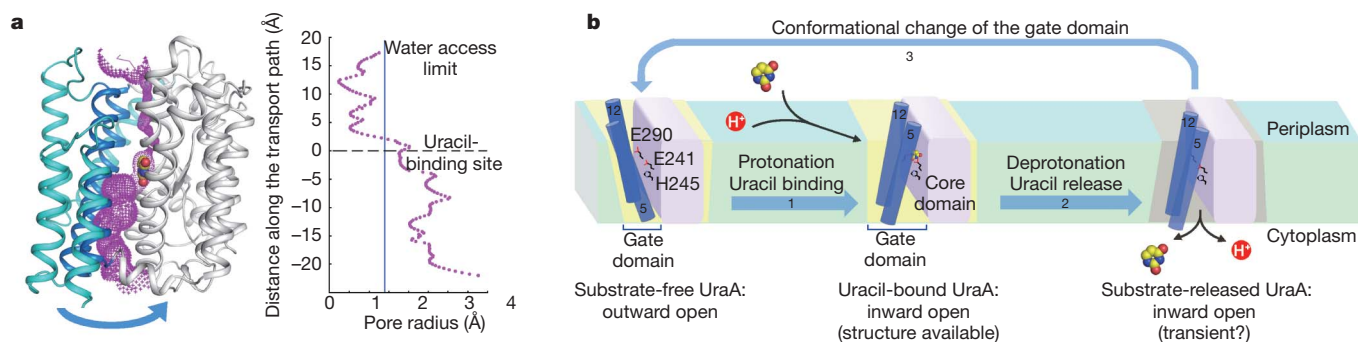

Figure 4 | Working model for the transport mechanism of UraA. a, Uracil-bound UraA is in an inward-open conformation. The van der Waals surface of UraA (purple) was calculated with the program HOLE[26], which revealed that the bound uracil is insulated from the periplasm but exposed to the cytoplasm. The pore radii along the potential transport path are tabulated on the right.

b, Working model to illustrate the putative transport mechanism of UraA. Glu 241, His 245 and Glu 290 are shown to emphasize their essential role in substrate recognition and proton translocation. For simplicity, only TM5 and TM12 are shown to represent the gate domain.

of the gate domain relative to the core domain around the bound uracil (Fig. 4a).

On the basis of structural and biochemical analyses, we propose a working model to explain proton-coupling and uracil symport by UraA (Fig. 4b). The default conformation of substrate-free UraA may be outward-open, with the two Glu residues deprotonated. The negative charges on the Glu residues make closure of the gate domain onto the core domain energetically unfavourable. On binding of uracil and H^+, at least one Glu residue is protonated and the gate domain undergoes a conformational change, taking an inward-open conformation as seen in the structure. A proton translocation is likely to occur from Glu to His 245 and onward into the cytoplasm. Deprotonation may cause local conformational changes around Glu 241, His 245 and Glu 290, leading to the dissociation of uracil. An inward-open, deprotonated conformation is likely to be transient and quickly reverts to the outward-open state. In the sodium symporters of the NAT family, the sodium ion, rather than the proton, is required to neutralize the conserved Glu and Asp residues and for substrate binding.

This model predicts that the core domain provides the molecular basis of substrate selectivity and proton/sodium translocation, whereas conformational changes of the gate domain allow transport of substrate. We recognize the speculative nature of this model, as many important questions remain unanswered and require experimental investigation. For example, we do not know the molar ratio between the proton/sodium ion and the substrate molecule during each transport cycle, nor do we know how the conserved Glu and His residues trigger conformational changes during protonation and deprotonation. Nonetheless, the structural and biochemical characterizations of UraA reported here provide an important framework for mechanistic understanding of the NAT family transporters.

METHODS SUMMARY

We generated all clones using a standard PCR-based cloning strategy. Wild-type and mutant UraA were expressed and purified to homogeneity. We grew crystals of wild-type UraA by the hanging-drop vapour diffusion method in the presence of 0.4% β-NG (Anatrace). All data were collected at SPring-8 beamline BL41XU and processed with the HKL2000 package[27]. The structure was determined by Hg-SIRAS using SOLVE[28] and refined with PHENIX[29]. We measured the binding affinity between uracil and purified recombinant UraA by scintillation proximity assay[17]. The cell-based uracil-uptake assay was modified on the basis of a published protocol[9].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Saier, M. H. Jr et al. Phylogenetic characterization of novel transport protein families revealed by genome analyses. *Biochim. Biophys. Acta* **1422,** 1–56 (1999).
2. Gournas, C., Papageorgiou, I. & Diallinas, G. The nucleobase-ascorbate transporter (NAT) family: genomics, evolution, structure-function relationships and physiological role. *Mol. Biosyst.* **4,** 404–416 (2008).
3. Tsukaguchi, H. et al. A family of mammalian Na^+-dependent L-ascorbic acid transporters. *Nature* **399,** 70–75 (1999).
4. Wang, Y. et al. Human vitamin C (L-ascorbic acid) transporter SVCT1. *Biochem. Biophys. Res. Commun.* **267,** 488–494 (2000).
5. Savini, I., Rossi, A., Pierro, C., Avigliano, L. & Catani, M. V. SVCT1 and SVCT2: key proteins for vitamin C uptake. *Amino Acids* **34,** 347–355 (2008).
6. Yamamoto, S. et al. Identification and functional characterization of the first nucleobase transporter in mammals: implication in the species difference in the intestinal absorption mechanism of nucleobases and their analogs between higher primates and other mammals. *J. Biol. Chem.* **285,** 6522–6531 (2010).
7. Diallinas, G. & Scazzocchio, C. A gene coding for the uric acid-xanthine permease of *Aspergillus nidulans*: inactivation cloning, characterization, and sequence of a cis-acting mutation. *Genetics* **122,** 341–350 (1989).
8. Karatza, P. & Frillingos, S. Cloning and functional characterization of two bacterial members of the NAT/NCS2 family in *Escherichia coli. Mol. Membr. Biol.* **22,** 251–261 (2005).
9. Andersen, P. S., Frees, D., Fast, R. & Mygind, B. Uracil uptake in *Escherichia coli* K-12: isolation of *uraA* mutants and cloning of the gene. *J. Bacteriol.* **177,** 2008–2013 (1995).
10. Koukaki, M. et al. The nucleobase-ascorbate transporter (NAT) signature motif in UapA defines the function of the purine translocation pathway. *J. Mol. Biol.* **350,** 499–513 (2005).
11. Karatza, P., Panos, P., Georgopoulou, E. & Frillingos, S. Cysteine-scanning analysis of the nucleobase-ascorbate transporter signature motif in YgfO permease of *Escherichia coli*: Gln-324 and Asn-325 are essential, and Ile-329–Val-339 form an α-helix. *J. Biol. Chem.* **281,** 39881–39890 (2006).
12. Georgopoulou, E., Mermelekas, G., Karena, E. & Frillingos, S. Purine substrate recognition by the nucleobase-ascorbate transporter signature motif in the YgfO xanthine permease: Asn-325 binds and Ala-323 senses substrate. *J. Biol. Chem.* **285,** 19422–19433 (2010).
13. de Koning, H. & Diallinas, G. Nucleobase transporters. *Mol. Membr. Biol.* **17,** 75–94 (2000).
14. Weyand, S. et al. Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science* **322,** 709–713 (2008).
15. Shimamura, T. et al. Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1. *Science* **328,** 470–473 (2010).
16. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. Crystal structure of a bacterial homologue of Na^+/Cl^-–dependent neurotransmitter transporters. *Nature* **437,** 215–223 (2005).
17. Quick, M. et al. Binding of an octylglucoside detergent molecule in the second substrate (S2) site of LeuT establishes an inhibitor-bound conformation. *Proc. Natl Acad. Sci. USA* **106,** 5563–5568 (2009).
18. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38,** W545–W549 (2010).
19. Screpanti, E. & Hunte, C. Discontinuous membrane helices in transport proteins and their correlation with function. *J. Struct. Biol.* **159,** 261–267 (2007).
20. Fu, D. et al. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **290,** 481–486 (2000).
21. Wang, Y. et al. Structure of the formate transporter FocA reveals a pentameric aquaporin-like channel. *Nature* **462,** 467–472 (2009).
22. Faham, S. et al. The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+/sugar symport. *Science* **321,** 810–814 (2008).
23. Ressl, S., Terwisscha van Scheltinga, A. C., Vonrhein, C., Ott, V. & Ziegler, C. Molecular basis of transport and regulation in the Na^+/betaine symporter BetP. *Nature* **458,** 47–52 (2009).
24. Fang, Y. et al. Structure of a prokaryotic virtual proton pump at 3.2 Å resolution. *Nature* **460,** 1040–1043 (2009).
25. Gao, X. et al. Mechanism of substrate recognition and transport by an amino acid antiporter. *Nature* **463,** 828–832 (2010).
26. Smart, O. S., Goodfellow, J. M. & Wallace, B. A. The pore dimensions of gramicidin A. *Biophys. J.* **65,** 2455–2460 (1993).
27. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
28. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55,** 849–861 (1999).
29. Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
30. DeLano, W. L. PyMOL Molecular Viewer ⟨http://www.pymol.org⟩ (2002).

Author Contributions F.L., S.L., Y.J., J.J., H.F., G.L. and N.Y. designed all experiments. F.L., S.L., Y.J., J.J., H.F., G.L., D.D., S.D. and X.Z. performed the experiments. All authors analysed the data. F.L., J.W. and N.Y. contributed to manuscript preparation. N.Y. wrote the manuscript.

Author Information The atomic coordinates of UraA have been deposited in the Protein Data Bank under accession code 3QE7. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.Y. (nyan@tsinghua.edu.cn).

METHODS

Protein preparation. The complementary DNA of full-length UraA from *E. coli* strain O157:H7 was subcloned into pET21b vector (Novagen). The UraA mutants were generated using two-step PCR and subcloned, overexpressed and purified in the same way as wild-type protein. Overexpression of UraA was induced in *E. coli* BL21(DE3) by 0.2 mM isopropyl-β-D-thiogalactoside (IPTG) when the cell density reached $D_{600\,nm}$ 1.5. To obtain the structure of uracil-bound UraA, uracil (Sigma) was added at 1 mM when cells were induced. Uracil (1 mM) was included in all the buffers during protein purification. After growth for 16 h at 22 °C, the cells were collected, resuspended in buffer containing 25 mM Tris-HCl, pH 8.0, and 150 mM NaCl, and disrupted using a French press with two passes at 10,000–15,000 p.s.i. Cell debris was removed by centrifugation at 27,000g for 10 min. The supernatant containing the membrane was collected and underwent ultracentrifugation at 150,000g for 1 h. The membrane fraction in the pellet was harvested and incubated with 1.5% (w/v) *n*-dodecyl-β-D-maltopyranoside (DDM, Anatrace) for 1 h at 4 °C. After another ultracentrifugation step at 150,000g for 30 min, the supernatant was collected and loaded on Ni^{2+}-nitrilotriacetate affinity resin (Ni-NTA, Qiagen). Subsequently, the resin was washed three times, each time with 10 ml buffer containing 25 mM Tris-HCl, pH 8.0, 150 mM NaCl, 30 mM imidazole and 0.02% DDM. The protein was eluted from the affinity resin with 10 ml wash buffer plus 250 mM imidazole. The proteins were concentrated to about 10 mg ml^{-1} before undergoing gel-filtration chromatography (Superdex-200 10/30, GE Healthcare) in buffer containing 25 mM Tris-HCl, pH 8.0, 150 mM NaCl and the indicated detergents. The peak fractions were collected for biochemical analyses or crystallization trials.

Crystallization. Crystals were grown at 18 °C by the hanging-drop vapour diffusion method. Crystals in the space group $P6_422$ were obtained for protein purified in the presence of 0.4% β-NG. Uracil (1 mM) was included during protein expression, purification and crystallization. Crystals appeared overnight in the buffer containing 25% PEG400, 100 mM MES-NaOH, pH 6.5, and 300 mM Li$_2$SO$_4$, and typically grew to form 50 μm × 50 μm × 100 μm hexagonal rods in about one month. The crystals diffracted X-rays beyond 2.9 Å at SPring-8 beamline BL41XU. Mercury derivatives were obtained by soaking the crystals for 24 h in the mother liquor containing 1 mg ml^{-1} (C$_2$H$_5$HgO)HPO$_2$. Both native and heavy-atom-derived crystals were directly flash frozen in a cold nitrogen stream at 100 K.

Data collection and structure determination. All the data were collected at SPring-8 beamline BL41XU and processed with the HKL2000 package[27]. Further processing was carried out with programs from the CCP4 suite[31]. Data collection statistics are summarized in Supplementary Table 1. The initial phases were obtained from the native $P6_422$ crystal and its Hg derivative by single isomorphous replacement with anomalous scattering (SIRAS) using the program SOLVE[28]. Figures of merit before and after density modification are 0.244 and 0.749, respectively. The real-space constraints were applied to the electron density map in DM[32]. The model was built manually in COOT[33] and the structure was refined with PHENIX[29]. All structure figures, including the calculation of surface electrostatic potential, were prepared with PYMOL[30].

Cell-based uracil-uptake assay. The *uraA*-deficient *E. coli* strain Keio Collection JW2482 (*F-*, Δ(*araD-araB*)567, Δ*lacZ4787*(::*rrnB-3*), λ-, Δ*uraA745::kan*, *rph-1*, Δ(*rhaD-rhaB*)568, *hsdR514*) used in this assay was purchased from National BioResource Project (Japan). Wild-type and UraA mutants were subcloned into pQLINK vector[34] with His$_6$ tag at the C terminus. The membrane expression levels of UraA variants were monitored by western blot using an antibody against the His tag. The amount of each UraA protein in the membrane fraction was estimated by comparing the intensity of the proteins in the membrane

fraction against a serial dilution of purified UraA with known concentrations on the same western blot, a protocol reported previously[35].

The cell-based uptake assay was performed following the published protocol with minor modifications[9]. *uraA*-deficient *E. coli* cells transformed with plasmids were grown in LB medium at 37 °C and induced with 50 μM IPTG when the cell density reached $D_{600\,nm}$ 1.5. Cells were collected after induction for 30 min, rinsed and resuspended to an adjusted $D_{600\,nm}$ 2.0 in AB medium (a modified minimal medium; see http://openwetware.org/wiki/AB_medium for the detailed recipe). After incubation in AB medium for 1 h, the cells were taken for the uracil-uptake assay.

For time-course experiments, [5,6-^{3}H]-uracil (4 Ci mmol^{-1}, American Radiolabelled Chemicals) was added at indicated concentrations. All the reactions were performed at 25 °C. At the indicated reaction time, an aliquot of cells was taken for filtration through 0.45-μm cellulose acetate filter (Sartorius). The filter membranes were immediately washed with 2 ml ice-cold AB medium, dried and taken for liquid scintillation counting. Control experiments were performed with cells transformed with pQLINK empty vector. For each data point, a parallel control experiment was performed and the control flux was subtracted before data fitting.

The time-course experiments showed that the accumulation of uracil was roughly linear within the first 30–60 s. Therefore, to determine the K_m and V_{max} of uracil uptake by wild-type UraA, the initial velocities were measured at 30 s. All experiments, including control experiments, were repeated at least three times, and the data were fitted to the Michaelis–Menten equation, $V = V_{max}[UraA]/(K_m + [UraA])$, in GRAPHPAD PRISM 5.0 Demo.

To compare the transport activity of the UraA variants, [5,6-^{3}H]-uracil (40 Ci mmol^{-1}) was applied at 0.25 μM and each reaction was allowed for 30 s. All the UraA variants were expressed and quantified following the same protocol as for wild-type UraA. The solution behaviour of the purified UraA variants was examined by size exclusion chromatography, which showed similar profiles to that of wild-type protein.

SPA-based binding assay. Scintillation proximity assay (SPA) beads were diluted to 2 mg ml^{-1} in 150 mM MES-NaOH, pH 6.4, 50 mM NaCl, 20% glycerol, 2 mM TCEP (Sigma) and 0.05% DDM. About 400 ng of purified, C-terminal His$_6$-tagged UraA protein was incubated with [5,6-^{3}H]-uracil at 4 °C for 2 h. The solution was then added to 100 μl SPA beads and incubated by vigorous shaking at 4 °C in the dark for 2 h. The mixture was then loaded into individual wells of 96-well plates. Scintillation was read in the SPA mode on a Wallac 1450 MicroBeta plate PMT counter. For the competition assay, the final concentration of [5,6-^{3}H]-uracil was kept constant at 0.17 μM while the concentration of non-labelled uracil was increased from 0 to 10 μM. To define the nonspecific binding activity, 400 mM imidazole was added to the wells to compete with His$_6$ for bead binding. All experiments were performed at least three times and data are presented as mean ± s.d. Nonspecific binding was subtracted from each data point. Data fitting was performed using GRAPHPAD PRISM 5.0 Demo.

31. Collaborative Computational Project, Number 4. The *CCP4* suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
32. Cowtan, K. DM: an automated procedure for phase improvement by density modification. *Joint CCP4 ESF-EACBM Newslett. Protein Crystallogr.* **31,** 34–38 (1994).
33. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
34. Scheich, C., Kummel, D., Soumailakakis, D., Heinemann, U. & Bussow, K. Vectors for co-expression of an unrestricted number of proteins. *Nucleic Acids Res.* **35,** e43 (2007).
35. Dang, S. *et al.* Structure of a fucose transporter in an outward-open conformation. *Nature* **467,** 734–738 (2010).

CAREERS

NATUREJOBS FORUM Employment advice and discussion at **go.nature.com/lm1x4t**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

IMAGES.COM/CORBIS

CONSULTING

Juggling act

US scientists who take on consultancy can gain experience, supplement their income — and face logistical challenges.

BY AMANDA LEIGH MASCARELLI

When Çağan Şekercioğlu had an interview for a position at the University of Utah in Salt Lake City in 2009, he faced a choice: play it safe and focus purely on his academic accomplishments and scientific publications, or openly discuss his environmental consultancy work as the founder and unpaid director of KuzeyDoğa, a non-governmental organization (NGO) based in Kars, Turkey.

Şekercioğlu suspected that his prospective colleagues might see his grass-roots conservation activity as a distraction from publishing high-impact papers and pursuing grants for his academic work. And traditional academic disciplines such as biology do not always value environmental work, he says. But Şekercioğlu decided to take the gamble and emphasize his voluntary work, hoping that the applied research involved would give him an edge. "I thought, 'Well, this is a big part of my identity'," he says. The gamble paid off: Şekercioğlu is now an assistant professor of biology at the University of Utah. "It's not enough for me just

to publish scientific papers. As a conservation biologist I must do actual conservation that leads to conservation of species," he says. Şekercioğlu now believes that the consulting experience helped him to land the job, and he has been pleasantly surprised by the support he has received from his academic colleagues.

Although some, like Şekercioğlu, may believe in a cause and a project enough to work on it for free, many US academics get paid for consultancy — and it can be lucrative. They often find that these side jobs allow them to branch out in ways that enrich their own research. And those who have chosen to consult full time say that it has given them diverse and satisfying opportunities. But being a consultant can be fraught with complexities — from the need to avoid actual or perceived conflicts of interest to managing one's time and taxes and navigating the delicate dynamics of relationships with faculty members and graduate students.

COMMON PRACTICE

Strictly speaking, consulting consists of paid work done on a contractual basis for an agreed period of time. But in practice it covers a wide range of activities, including voluntary work. Many scientists, for example, sit on scientific boards or on the boards of directors of large companies or organizations. Consultancy pay varies widely, and usually generates anywhere from US$75 to $300 per hour. For researchers who regularly act as consultants, the extra income can be as much as 30% of their main salaries.

Academics often rent out their expertise under consultancy arrangements. During the Deepwater Horizon oil spill in 2010, numerous researchers were hired by BP (see *Nature* **466,** 538; 2010) and the US National Oceanic and Atmospheric Administration (NOAA) to collect samples and to analyse data and environmental effects. Academic researchers are often called on to act as expert witnesses in trials, for example, to interpret forensic evidence or to assess damage to health or the environment from a chemical accident.

Consulting opportunities span all scientific disciplines, but certain fields and institutions receive more funding from industry than others, which can lead to more consulting opportunities, says John Humphrey, head of the geology and geological engineering department at the Colorado School of Mines in Golden. "The funding automatically puts us in contact with people in industry; they see what we do, and ▶

we're working with them anyway," he says. Companies needing an expert interpretation of data, or wanting to use a technique for which they don't have the necessary analytical equipment, will often contact scientists that they know for a consultation, adds Humphrey.

SEEKING AND FINDING

Those interested in seeking out consultancy work have several avenues to explore. For example, the American Chemical Society and the American Institute of Chemical Engineers sponsor an online forum called the Chemical Consultants Network (chemconsultants. org), a non-profit organization that brings together some 500 chemists and chemical engineers from the United States and elsewhere, allowing them to publicize their consultancy services and expertise. A similar network, the CECON Group (cecon.com), acts as a broker for scientists, engineers and expert witnesses. And researchers worldwide have become involved with consultancy through innovation challenges such as those hosted online by InnoCentive and IdeaConnection (see *Nature* **469**, 433–435; 2011).

Academic scientists who consult on a regular basis say that the jobs often develop through word of mouth and networking. But the practice can be more institutionalized. The Florida Institute of Technology in Melbourne has its own consultancy department, Florida Tech Consulting, which acts as a sort of "clearing house for consulting expertise", says Joel Olson, a chemist at Florida Tech. A company in need of a consultant with specific expertise contacts the department, which provides a list of candidates.

Shelley Johnson, associate director of Florida Tech Consulting, estimates that about a quarter of Florida Tech's faculty members sometimes consult, and the number is growing. Johnson says they recognize the appealing prospect of earning significantly more than their regular salary.

"They're quick little projects where I get to learn about some new area of science."

Consultancy often requires a skill set beyond typical lab or field research. For example, Daniel Cooper, president of Cooper Ecological Monitoring, an environmental consultancy in Los Angeles, California, sometimes finds himself accessing antiquated field notes at museums and surveying railway reports. He occasionally advises clients on habitat restoration projects, which might entail looking at old maps and photos to get a sense of the area's native vegetation and topography. He has also designed an 'eco audit' to help a

coffee-importing company verify that the coffee it is buying is grown sustainably, and is helping a city to clear brush in fire-prone areas without disturbing sensitive plants and nesting birds.

The business side of things is important too: operations such as Cooper's require that the consultant does everything from drawing up proposals and estimates of the scope of work to invoicing the clients and applying for renewal of permits to survey endangered species. Cooper also often has to decipher archaic names of plant species and communicate with non-English-speaking maintenance crews to convey such information as why it is important

Çağan Şekercioğlu, a biologist at the University of Utah, also runs two environmental organizations in Turkey.

not to cut down branches that are home to nesting birds. "You're really a liaison between the natural world and the 'realities' of getting work done," says Cooper.

SAFE PRACTICE

Liability insurance is needed to protect consultants from lawsuits, says Cooper. Responsibility for it varies: Florida Tech's contract automatically shields scientists, says Olson. Cooper notes that two of his recent projects required that he get his own professional liability insurance, which he estimates will cost him an extra $1,000 a year (although one client offered to cover the cost). Registering as a 'limited liability company' — or one's country's equivalent — confers some protection. John Newport, a full-time chemistry consultant with Chemventive in Chadds Ford, Pennsylvania, recommends incorporating liability limits up front in the contract with the client. Some

part-time consultants, however, say they have never purchased insurance.

Unconventional tasks and protections are not the only challenge for the consultant scientist. The work can also clash with the culture of a researcher's home institution, creating strain among colleagues. One oceanographer, who prefers to remain anonymous owing to concerns that some of his colleagues would not approve of his consulting activities, notes that in academia, consultancy can sometimes create inequities with other faculty members. "In the system I'm in, the university is set up to be fair in terms of promotions and pay," he says. "When people are doing consulting, which can be a significant source of income, it's not something we openly talk about. Many academics believe in the purity of academic pursuit, and their perception is that consulting is tainted." This researcher, who began working as a consultant for NOAA during the 2010 oil spill, emphasizes that faculty members must avoid letting consultancy interfere with the work of their graduate students or postdoctoral researchers. For instance, data collected during consultancy for the government or a company are often kept private, and researchers must make sure that these data aren't vital to anyone else's work. Companies or businesses should not have the power to interfere with publication of a student or postdoc's research, he says. "If somebody's career could potentially be impacted in a negative way, then there's a firm line."

Most universities require that faculty members disclose, in writing, the nature of their consultancy work and the time they devote to it. At the Colorado School of Mines, faculty members are allowed to spend up to eight hours a week on consultancy work. "You don't want to be in a position where someone is questioning whether you're getting your principal job done," says Humphrey. He estimates that up to a third of faculty members in the department of geology and geological engineering do some form of consultancy.

When Şekercioğlu, who now runs two Turkish environmental NGOs, wraps up his university day, he begins his consultancy work — coordinating with his staff, collaborators, students, volunteers and other partners in Turkey. "Often you'll find me on Skype or the phone after midnight," says Şekercioğlu.

Humphrey has worked on several international consultancy assignments in central Asia, usually helping multinational oil companies to understand the geology of oil reservoirs. But juggling consultancy and a full-time faculty position can be challenging. "You have to strike the balance between doing what you do on a daily basis and trying to squeeze something else in," he says.

T. TAZEGUN

As head of the department, Humphrey has to sign off on all faculty members' consultancy jobs. He cautions that researchers must maintain clear boundaries by, for example, not mingling consultancy work with university travel unless it is explicitly approved. He also notes that in some countries, including the United States, taxes are usually not taken out of payment up front, so consultants must plan accordingly and set taxes aside — or risk penalties. All scientists who act as consultants, whether full or part time, must grapple with the sometimes-elusive concept of doing unbiased science for a paying client who has a vested interest in the outcome. "You should question yourself on every job," says Şekercioğlu. "When you're working for industry and big government, it's hard to resist that pressure sometimes." Working for free helps Şekercioğlu to stick to his principles, he says. "I can stand my ground. But consulting is often not black and white. It's grey."

> **"When people are doing consulting, it's not something we talk openly about. Many academics believe in the purity of academic pursuit."**

"Maintaining scientific neutrality can sometimes require significant professional discipline," says Olson. "I always strive to do good science, but science is not a cut-and-dried field of work. Results can be interpreted in different ways." One of Olson's clients, a consortium of chemical companies, paid him to evaluate some scientific papers and investigate whether the methodology in them was sound; he found egregious flaws. Olson says that being paid didn't influence the quality of his work, but it was always in the back of his mind. "I know who's paying me and I know what they want to see," he says. Consultants should make sure clients understand that they intend to report the scientific results, whatever they may be, says Olson. "And if the client is legitimate, they'll not only accept robust science, but they'll be enthusiastic about it," he adds. "I wouldn't work for someone otherwise."

Even with all the caveats, scientists who act as consultants often find it fulfilling. The anonymous oceanographer says that one of his consultancy projects paid him to do a literature review that was highly valuable for his own research. "For me, it's kind of a dream thing," he says. Dabbling in diverse areas is part of the reward. "They're quick little projects," says Olson, "where I get to learn about some new area of science." ∎

Amanda Leigh Mascarelli *is a freelance writer in Denver, Colorado.*

TURNING POINT
Giovanna Tinetti

Giovanna Tinetti, a planetary scientist at University College London, learned in February that her team's proposal to lead the £400-million (US$642-million) Exoplanet Characterisation Observatory (EChO) mission to search for life on other planets will be backed by the European Space Agency (ESA).

You started off as a theoretical particle physicist. How did you make the leap to searching for life on exoplanets?
I was pursuing my PhD in theoretical physics at the University of Turin in Italy, yet I was increasingly interested in working on something for which experiments were the driving force. So I started to look at other possibilities. In 1998, NASA started a virtual Astrobiology Institute to prepare ambitious experiments looking for life and habitable planets in the Universe. As I learned more, I decided that it was a great move for me. I was intrigued by ideas about Gaia and the notion that abiotic planets and living organisms grow together. I began my adventure with exoplanets as a postdoc at NASA's Jet Propulsion Laboratory in Pasadena, California.

Characterize your early career choices.
I made a series of decisions not to follow the 'safe' route. First, I turned down a well-paying job, instead choosing to finish my PhD. And just deciding to begin this search for exoplanets meant making a big bet with my career. Back then, only a few extrasolar planets had been discovered, and nobody knew whether they would prove viable for life. Several people told me that this could be a path that leads nowhere. I just felt that I had one life, and if I didn't try, I would regret it. Luckily, the field has been successful, so I made the right choices.

How have your past experiences shaped how you approached this space mission?
During my postdoc, a number of extrasolar planets were discovered, and attention shifted to a dedicated experiment to probe exoplanet atmospheres. It ultimately became a joint project, combining NASA's Terrestrial Planet Finder (TPF) and ESA's Darwin. But technological challenges and budget hurdles kept pushing the launch date further and further back. At the same time, we discovered that we could look for atmospheres that signalled potentially habitable worlds by monitoring planets as they move in front of their stars. This 'transit spectroscopy' technique proved successful, and we started to think that, rather than pursuing a big experiment, we should use the technique with existing telescopes, such as Hubble or Spitzer. So the TPF/Darwin mission was shelved. But the process of preparing for a big experiment was informative.

Did that help you achieve success with EChO?
Absolutely. Using the transit technique and existing technology, we can now make the most of a dedicated exoplanet mission. The most important technology — including a 1.2-metre telescope and spectrograph — already exists. That was important when submitting a proposal for a launch in 2020.

Are you confident EChO will launch in 2020?
I plan to work hard to ensure its success. It looks like a long lead time but it's really not. We have to go through ESA's assessment phase, to judge whether the project is doable — so we have to be ready to answer any question relating to the science.

Are you taking steps to safeguard your career in case EChO is derailed?
The preparation for EChO relates to my everyday research. My team continues to work hard on observing planets from the ground and from the Hubble and Spitzer telescopes. When you are involved in a space mission, you can't bet on its success until you see the satellite in orbit. That said, we are doing our best to create a mission that benefits extrasolar-planet research.

What is the most important thing you've done for your career?
I can say that it was a good idea to spend one year of my life pursuing this space mission because, in the end, we were selected. You have to believe in something — unfortunately, that doesn't necessarily mean it will happen. ∎

INTERVIEW BY VIRGINIA GEWIN

STEVE SEPP, TASTY! TASTY!

No mean feast.

BY MATTHEW SANBORN SMITH

We didn't eat Steve on a Tuesday, which I think was one of the things that made him special.

Steve Sepp — he was one of those all-time losers you talk about to make your own life seem bearable. Lived in a mobile home that hadn't run in a year and a half (that was what he called his car, anyway); couldn't get work; divorced; running from everybody. You didn't want to hang with the guy, because you didn't want to catch whatever he had, you know what I mean? But I felt bad for him and other people did too. The wife would bring him a loaf of bread she baked or some corn muffins or whatever, usually Tuesday nights when we were on our way to the hospital.

This one night though, Steve's sick as a dog, doubled over in pain. He took it until he couldn't take it anymore, which wasn't long really, 'cause Steve was a major wuss on top of all his other problems. Mira and I, we hustle him over to the ER. Nurse asks a couple questions, presses her hand against his belly button. "Might be your appendix," she says, and they grab a wheelchair and whisk him off.

Of course, of all the people we know, Steve was the guy who gets appendicitis. That's his luck. "Well, at least he doesn't have to wait," I says to Mira. She slaps my arm and we head to the south end of the building where we wanted to go in the first place.

The Baker Street Hospital is the place to go on Tuesday nights if you aren't already sick. Tuesday's usually their biggest outpatient day and the Outpatient Diner is what draws people from all over the neighbourhood and beyond. All those little bits and pieces they cut off of people during the day, they wind up on the menu. You know, stew, pulled-meat sandwiches, what have you. If you're an outpatient, you get a freebie after your surgery. Most of 'em want a bite of themselves. You could call it cannibalism I suppose, but everyone who was on that menu was up and walking around (or wheeling around). The hospital has the waivers. It's all on the up and up.

Not having any insurance, Steve signed the waivers without a blink because it defrayed some of his costs. Not a lot, but every little bit helps. He was in the hole as it was. If he lived through this, it would kill him.

So, like I say, Tuesday nights are hopping at the diner. It got so that regulars like me and Mira couldn't even get in. We're getting bumped by celebrities, for chrissakes, like the guy who does Sharona LaHaye's hair and that pip who sells Maxi all her pet supplies.

So we give up and go back on Wednesday, hoping maybe to get a discount on day-old spleen soup or something, I don't know what we were thinking, but we were glad we did. 'Scuse me, my mouth is watering just remembering it.

We walk in not knowing what's what and the place is nuts. Jimmy Palbro runs up to us before we even sit down, and says, he says, "Al, Mira, you guys gotta try this!" He's got a couple of biscuits with some gravy on 'em like it's the south or somethin', only it's brown gravy. Anyway we taste 'em and me and Mira, our eyes bug out.

"What is this?" I ask him. And he says: "It's Steve Sepp gravy! Steve from down the street? They musta brought him in for somethin'."

"Yeah," I says, "Mira and me. We brought him in last night."

"You hear that, everybody?" Jimmy Palbro yells. "Al and Mira are the ones who brought Steve in!" And they're hollerin' and applaudin' like we're the ones who made him sick.

Generally, you go down to the diner, you meet your buddies, the regulars, and you talk about the Sox or politics or whatnot, but this Wednesday, and I remember it like it was yesterday, this Wednesday all we can talk about is Steve. Holy Jesus, you should have tasted him on the mashed potatoes! When they wheeled him in for his share, the chef actually announced him and I tell you, we stood up and clapped and whistled louder than before. He coulda run for friggin' mayor that day. We would've carried him to his office on our shoulders. And he knew it too.

He's sittin' there in his wheelchair, lookin' skin-boney, like he always has, miserable as hell, and when he really gets it, that we're clapping for him, he sits up straighter. He even smiles. Well, with what teeth he has.

Somebody hands him a little plate of potatoes covered in himself. That was like a feast to Steve, even on a good day. The whole place gets quiet, and we're just standing there, watching him. The nurse gives him a spoon and he gets a little scoop, he's shakin' a little, and he puts it in his mouth and you can see his eyes get really wide. And he works his mouth a couple of times, but you don't really gotta chew it, you know that hospital stuff is basically mush, but he works it and he swallows and he goes to everybody: "I'm really good!"

And the whole house goes crazy! ∎

Matthew Sanborn Smith's work has appeared at Tor.com and in Chizine, GUD Magazine and the StarShipSofa podcast. He podcasts at http://bewarethehairymango.com.

MENU

CORRECTIONS & AMENDMENTS

Mediator and cohesin connect gene expression and chromatin architecture

Michael H. Kagey, Jamie J. Newman, Steve Bilodeau, Ye Zhan,
David A. Orlando, Nynke L. van Berkum,
Christopher C. Ebmeier, Jesse Goossens, Peter B. Rahl,
Stuart S. Levine, Dylan J. Taatjes, Job Dekker
& Richard A. Young

It is important to note that components of the cohesin and mediator complexes, including *Smc1a*, *Nipbl*, *Rad21* and *Med10*, were identified in previous RNA interference screens for maintenance of embryonic stem cells[1–3].

1. Hu, G. *et al.* A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev.* **23,** 837–848 (2009).
2. Fazzio, T. G., Huff, J. T. & Panning, B. An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* **134,** 162–174 (2008).
3. Ding, L. *et al.* A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. *Cell Stem Cell* **4,** 403–415 (2009).

CORRECTIONS & AMENDMENTS

Holocene Southern Ocean surface temperature variability west of the Antarctic Peninsula

A. E. Shevenell, A. E. Ingalls, E. W. Domack & C. Kelly

On page 3 of the PDF and print version of this Letter, column two, at the end of the fourth paragraph, the last line inadvertently read "occurred between 1,600 and 500 kyr BP", but this should have said "occurred between 1,600 and 500 years BP". This has been corrected in the HTML version of the manuscript.

CORRECTIONS & AMENDMENTS

H2AX prevents CtIP-mediated DNA end resection and aberrant repair in G1-phase lymphocytes

Beth A. Helmink, Anthony T. Tubbs, Yair Dorsett,
Jeffrey J. Bednarski, Laura M. Walker, Zhihui Feng,
Girdhar G. Sharma, Peter J. McKinnon, Junran Zhang,
Craig H. Bassing & Barry P. Sleckman

In Figure 2b of this Letter, the '+' and '−' labels for the iATM lanes of the Southern blot analysis were inadvertently switched round. The corrected Figure 2b is shown below.